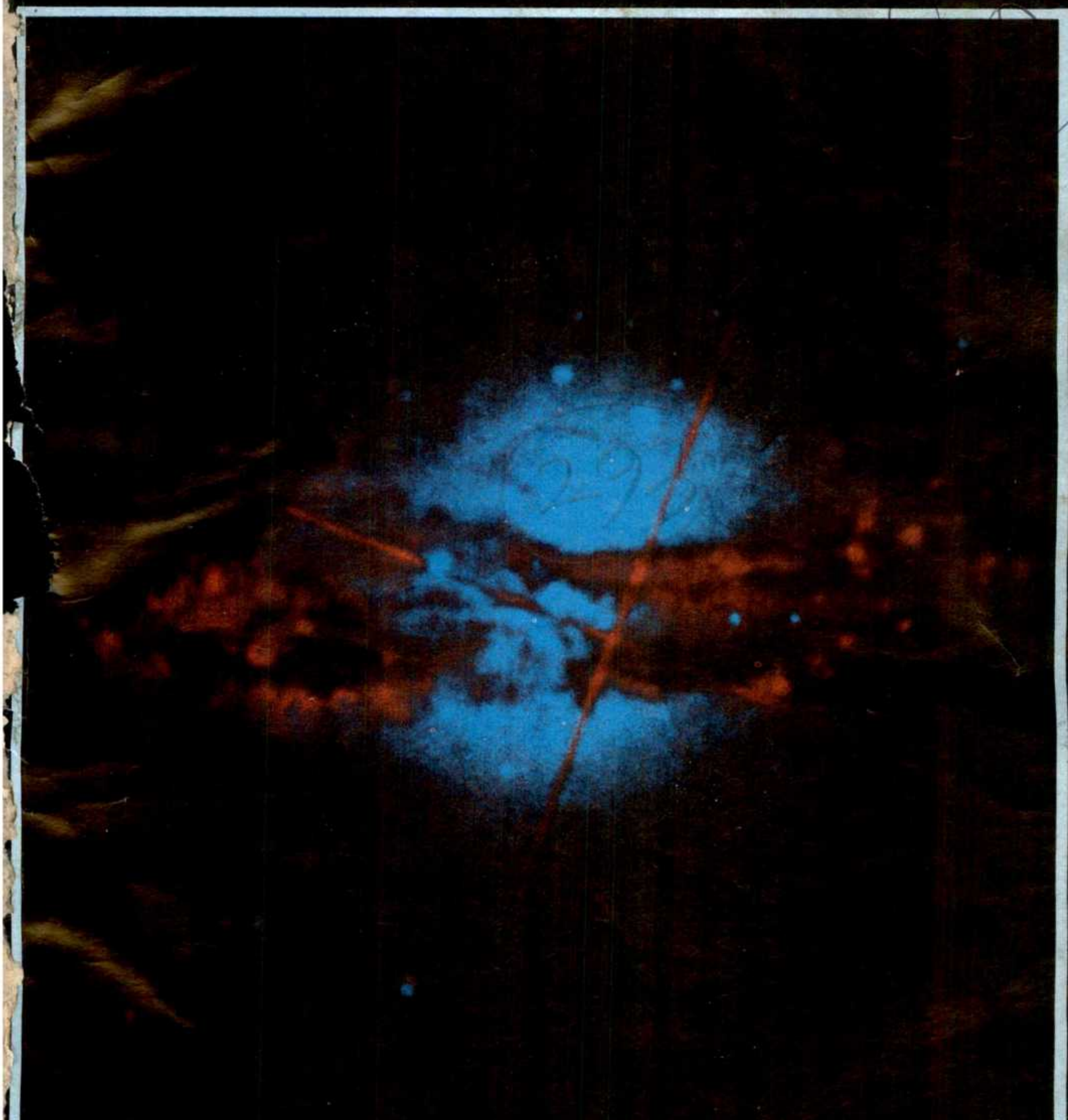
nature

Vol 297 No 5861 6–12 May 1982 £1.60 $4.00



EVOLUTION OF THE GALAXY CENTAURUS A

14-11-83

nature

. 297 No. 5861 6 May 1982

over
false-colour image of the powerful radio source en A. Observations reported on page 38 suggest that his galaxy is a spiral in which the disk formed much ter than the bulge.

ture® (ISSN 0028-0836) is published weekly on Thursday, except the last week in December, by Macmillan Journals Ltd and includes the Directory of Biologicals (mailed in December) and Annual ex (mailed in February). Annual subscription for USA and Canada US $198.50 (for subscription prices elsewhere, see next page). Orders (with remittance) and change of address labels to: Macmillan rnals Ltd, Brunel Rd, Basingstoke RG21 2XS, UK. Second class postage paid at New York, NY10010 and additional mailing offices. US Postmaster send form 3579 to: *Nature*, 15 East 26 eet, New York, NY 10010.

028-0836/82/18000C-02$01.00

Contents continued overleaf

GUIDE TO AUTHORS

Authors should be aware of the diversity of *Nature*'s readership and should strive to be as widely understood as possible.

Review articles should be accessible to the whole readership. Most are commissioned, but unsolicited reviews are welcome (in which case prior consultation with the office is desirable).

Scientific articles are research reports whose conclusions are of general interest or which represent substantial advances of understanding. The text should not exceed 3,000 words and six displayed items (figures plus tables). The article should include an abstract of about 50 words.

Letters to *Nature* are ordinarily 1,000 words long with no more than four displayed items. The first paragraph (not exceeding 150 words) should say what the letter is about, why the study it reports was undertaken and what the conclusions are.

Matters arising are brief comments (up to 500 words) on articles and letters recently published in *Nature*. The originator of a Matters Arising contribution should initially send his manuscript to the author of the original paper and both parties should, wherever possible, agree on what is to be submitted.

Manuscripts may be submitted either to London or Washington (decisions being made only in London). Manuscripts should be typed (double spacing) on one side of the paper only. Three copies are required, each accompanied by copies of lettered artwork. No title should exceed 80 characters in length. Reference lists, figure legends, etc. should be on separate sheets, all of which should be numbered. Abbreviations, symbols, units, etc. should be identified on one copy of the manuscript at their first appearance.

References should appear sequentially indicated by superscripts in the text and should be abbreviated according to the *World List of Scientific Periodicals*, fourth edition (Butterworth 1963–65). The first and last page numbers of each reference should be cited. References to books should clearly indicate the publisher and the date and place of publication. Unpublished articles should not be formally referred to unless accepted or submitted for publication, but may be mentioned in the text.

Each piece of artwork should be clearly marked with the author's name and the figure number. Original artwork should be unlettered. Suggestions for cover illustrations are welcome. Original artwork (and one copy of the manuscript) will be returned when manuscripts cannot be published.

Requests for permission to reproduce material from Nature *should be accompanied by a self-addressed (and, in the case of the UK and USA, stamped) envelope.*

Annual Subscription including Index (51 issues)

UK & Eire		£85
USA & Canada		US$198.50
Belgium	Airspeed only	BF7450
West Germany	Airspeed only	DM450
Netherlands	Airspeed only	G505
Switzerland	Airspeed only	SF395
Rest of Europe	Airspeed only	£100
Rest of World	Surface	£100
Rest of World*	Airmail	£150

**(not USA, Canada & Europe)*

Personal subscription rates are available in some countries to subscribers paying by personal cheque or credit card.

Details can be obtained from:
UK: *Nature* Promotion Department, Canada Road, Byfleet, Surrey KT14 7JL.
Telephone: Jonathan Earl: Byfleet (09323) 41459
USA: *Nature*, 15 East 26 Street, New York, NY 10010
Credit card orders only (in USA & Canada):
Call toll-free: (800) 824-7888 (Operator 246)
In California: (800) 857-7777 (Operator 246)

All other circulation enquiries concerning existing subscriptions please contact:
Nature Circulation Department, Macmillan Journals Ltd., Brunel Road, Basingstoke, Hants RG21 2XS, UK
Telephone: (0256) 29242 Telex: 858493

Back issues: (Post-paid) UK, £2.50; USA & Canada, US$6.00 (surface), US$9.00 (air); Rest of World, £3.00 (surface), £4.00 (air).

***Nature* Binders:**
Single Binders: UK, £4.00; Rest of World, $10.00
Set(s) of 3 Binders: UK, £10.00; Rest of World, 25.00

***Nature* First Issue Facsimile:**
UK, 75p; Rest of World (surface), $1.50, (air), £2.00

***Nature* Annual Indexes (1971-1980):**
Price (post-paid):
UK, £5.00 each; Rest of World, $10.00

***Nature* Directory of Biologicals:**
(Hardcover edition)
UK, £22.50; Rest of World, $45.00

***Nature* Wallcharts:**
How Britain Runs its Science; How the US Runs its Science; UK, £3.50; USA, $7.95 each

***Nature* in Microform:**
For Information:
UK: UMI, 18 Bedford Row, London WC2R 4EJ
USA: 300 North Zeeb Road, Ann Arbor, Mi 48106

Orders (with remittance) to:
Nature, Macmillan Journals Ltd.,
Brunel Road, Basingstoke, Hants RG21 2XS, UK
Tel: (0256) 29242 Telex: 858493

MACMILLAN JOURNALS

EDITORIAL OFFICES

London
4 Little Essex Street, WC2R 3LF
Telephone: (01) 836 6633 Telex: 262024

Editor: John Maddox

Deputy Editor: Peter Newmark

Editorial Staff
Alun Anderson, Geoffrey North
Philip Campbell, Peta Pickering
Isobel Collins, Judy Redfearn
Tim Lincoln, Miranda Robertson
David Millar, Robert Walgate
Naomi Molson, Charles Wenz
Sara Nash, Nigel Williams

Washington
Washington Editor: Deborah Shapley
991 National Press Building, DC 20045
Telephone: (202) 737-2355 Telex: 64280

Publisher: Elizabeth Hughes
Marketing Director: Ray Barker
International Advertising Manager: Andy Sutherland
Features Advertising Manager: Marion Delaney
Classified Sales Manager: Jean Neville
Telephone: (01) 240 1101
Promotion Manager: Jonathan Earl

New York
15 East 26 Street, New York, NY 10010
Telephone: (212) 689-5900
American Publisher: Robert Ubell
American Advertising Manager: Henry Dale
Marketing Services Manager: Sheila Kane
USA Classified Advertising: Cathy Moore

Canadian Display and Classified
Peter Drake, 32 Front Street West,
200 Toronto, Ontario M5J IC5
(416) 364-1623

Display Advertising Representatives

USA
Nature New York Office—(212) 689-5900
Roy McDonald Associates, Inc. Dallas—(214) 941-4461
Roy McDonald Associates, Inc. Houston—(713) 988-5005
Didier & Broderick, Inc. Chicago—(312) 498-4520
Jobson/Jordan/Harrison/Schulz, Inc. San Francisco (415) 392-6794
Los Angeles & Pasadena (213) 796-9200
Daniel Adams Associates, Inc. Philadelphia—(215) 353-6191
CEL Associates Boston—(617) 848-9306
Brinker & Brinker Fort Lauderdale (305) 771-0064

West Germany and Austria
Franz Schrecklinger
TMW Top Media Werbegesellschaft GmbH
Frankfurt (611) 726046

Holland and Belgium
Kevin Ward
G. Arnold Teesing BV
Amsterdam (020) 2636 15

Switzerland
Verena Hugi
Agentur Iff AG
Schaffhausen (053) 45821

Scandinavia
Andrew Karnig
Andrew Karnig & Associates AB
Stockholm (08) 51 68 70

Japan
Mashy Yoshikawa
Orient Echo Inc
Tokyo 541-4923

Nature®
ISSN 0028-0836
Registered as a newspaper at the British Post Office

Vol. 297 No. 5861 6 May 1982

RP123

nature

Vol. 297 No. 5862 13 May 1982

NEWS

CORRESPONDENCE

NEWS AND VIEWS

BOOK REVIEWS

Cover
Steel reinforcement corrosion in concrete — see page 109.

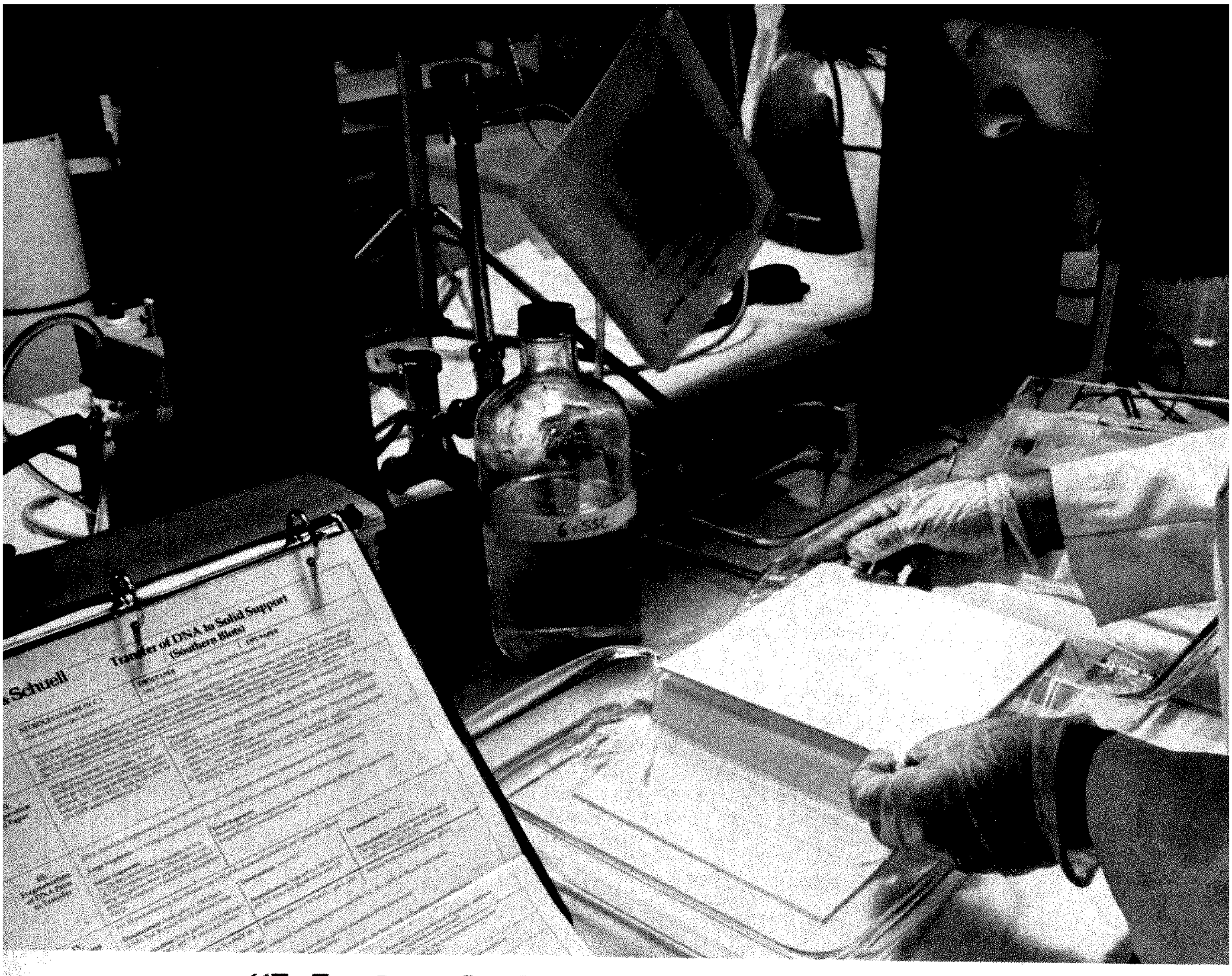
ARTICLES

LETTERS TO NATURE

Nature® (ISSN 0028-0836) is published weekly on Thursday, except the last week in December, by Macmillan Journals Ltd and includes the Directory of Biologicals (mailed in December) and Annual Index (mailed in February). Annual subscription for USA and Canada US $198.50 (for subscription prices elsewhere, see next page). Orders (with remittance) and change of address labels to: Macmillan Journals Ltd, Brunel Rd, Basingstoke RG21 2XS, UK. Second class postage paid at New York, NY10010 and additional mailing offices. US Postmaster send form 3579 to: *Nature*, 15 East 26 Street, New York, NY 10010.

0028-0836/82/19000C-02$01.00

Contents continued overleaf



"Materials and Methods" S&S sets the standard

Key reasons researchers prefer S&S nitrocellulose, such as BA85, are its versatility and the fact that it is the material upon which the *entire* hybridization field is based. The first published report, in 1963, specified S&S nitrocellulose for filter hybridization. S&S nitrocellulose is still the accepted standard.

Now S&S nitrocellulose is used extensively in Southern (DNA) transfers, RNA transfers, bi-directional transfers, electrophoretic transfers, protein transfers, as well as plaque and colony hybridization studies. No other transfer medium has stood the test of time so well.

Another important factor behind the preeminence of S&S in the genetics field is the S&S policy of *total* commitment to the user. S&S understands the applications of nitrocellulose and provides technical assistance, comprehensive applications literature and the "Sequences" newsletters.

Photographed at the Institute of Cancer Research, Columbia University, Dr. Philip Feigelson's laboratory.

Moreover, 18 years of experience as a manufacturer of nitrocellulose enables S&S to guarantee product quality, lot after lot.

Publication of the basic transfer methods (as shown) is the latest example of S&S total commitment. Written by a team of leading molecular biologists, this 8-page guide details DNA, RNA and protein transfer techniques using nitrocellulose and Transa-Bind™ diazotized media. It is available free upon request from S&S, manufacturer of the recognized standard.

Schleicher & Schuell

Schleicher & Schuell, Inc.
Keene, New Hampshire 03431
(603) 352-3810

Schleicher & Schüll GmbH, D-3354, Dassel, West Germany

Schleicher & Schüll AG, CH-8714 Feldbach ZH, Switzerland

Circle No.07 on Reader Service Card.

MISCELLANY

GUIDE TO AUTHORS

Authors should be aware of the diversity of *Nature*'s readership and should strive to be as widely understood as possible.

Review articles should be accessible to the whole readership. Most are commissioned, but unsolicited reviews are welcome (in which case prior consultation with the office is desirable).

Scientific articles are research reports whose conclusions are of general interest or which represent substantial advances of understanding. The text should not exceed 3,000 words and six displayed items (figures plus tables). The article should include an abstract of about 50 words.

Letters to *Nature* are ordinarily 1,000 words long with no more than four displayed items. The first paragraph (not exceeding 150 words) should say what the letter is about, why the study it reports was undertaken and what the conclusions are.

Matters arising are brief comments (up to 500 words) on articles and letters recently published in *Nature*. The originator of a Matters Arising contribution should initially send his manuscript to the author of the original paper and both parties should, wherever possible, agree on what is to be submitted.

Manuscripts may be submitted either to London or Washington (decisions being made only in London). Manuscripts should be typed (double spacing) on one side of the paper only. Three copies are required, each accompanied by copies of lettered artwork. No title should exceed 80 characters in length. Reference lists, figure legends, etc. should be on separate sheets, all of which should be numbered. Abbreviations, symbols, units, etc. should be identified on one copy of the manuscript at their first appearance.

References should appear sequentially indicated by superscripts in the text and should be abbreviated according to the *World List of Scientific Periodicals*, fourth edition (Butterworth 1963–65). The first and last page numbers of each reference should be cited. References to books should clearly indicate the publisher and the date and place of publication. Unpublished articles should not be formally referred to unless accepted or submitted for publication, but may be mentioned in the text.

Each piece of artwork should be clearly marked with the author's name and the figure number. Original artwork should be unlettered. Suggestions for cover illustrations are welcome. Original artwork (and one copy of the manuscript) will be returned when manuscripts cannot be published.

Requests for permission to reproduce material from Nature *should be accompanied by a self-addressed (and, in the case of the UK and USA, stamped) envelope.*

Annual Subscription including Index (51 issues)

UK & Eire		£85
USA & Canada		US$198.50
Belgium	Airspeed only	BF7450
West Germany	Airspeed only	DM450
Netherlands	Airspeed only	G505
Switzerland	Airspeed only	SF395
Rest of Europe	Airspeed only	£100
Rest of World	Surface	£100
Rest of World*	Airmail	£150

**(not USA, Canada & Europe)*

Personal subscription rates are available in some countries to subscribers paying by personal cheque or credit card.

Details can be obtained from:
UK: *Nature* Promotion Department,
Canada Road, Byfleet, Surrey KT14 7JL.
Telephone: Jonathan Earl: Byfleet (09323) 41459
USA: *Nature*, 15 East 26 Street, New York, NY 10010
Credit card orders only (in USA & Canada):
Call toll-free: (800) 824-7888 (Operator 246)
In California: (800) 852-7777 (Operator 246)

All other circulation enquiries concerning existing subscriptions please contact:
Nature Circulation Department, Macmillan Journals Ltd., Brunel Road, Basingstoke, Hants RG21 2XS, UK
Telephone: (0256) 29242 Telex: 858493

Back issues: (Post-paid) UK, £2.50; USA & Canada, US$6.00 (surface), US$9.00 (air); Rest of World, £3.00 (surface), £4.00 (air).

***Nature* Binders:**
Single Binders: UK, £4.00; Rest of World, $10.00
Set(s) of 3 Binders: UK, £10.00; Rest of World, 25.00

***Nature* First Issue Facsimile:**
UK, 75p; Rest of World (surface), $1.50, (air), £2.00

***Nature* Annual Indexes (1971-1980):**
Price (post-paid):
UK, £5.00 each; Rest of World, $10.00

***Nature* Directory of Biologicals:**
(Hardcover edition)
UK, £22.50; Rest of World, $45.00

***Nature* Wallcharts:**
How Britain Runs its Science; How the US Runs its Science; UK, £3.50; USA, $7.95 each

***Nature* in Microform:**
For Information:
UK: UMI, 18 Bedford Row, London WC2R 4EJ
USA: 300 North Zeeb Road, Ann Arbor, Mi 48106

Orders (with remittance) to:
Nature, Macmillan Journals Ltd.,
Brunel Road, Basingstoke, Hants RG21 2XS, UK
Tel: (0256) 29242 Telex: 858493

MACMILLAN JOURNALS

MEETING THE CHALLENGE OF

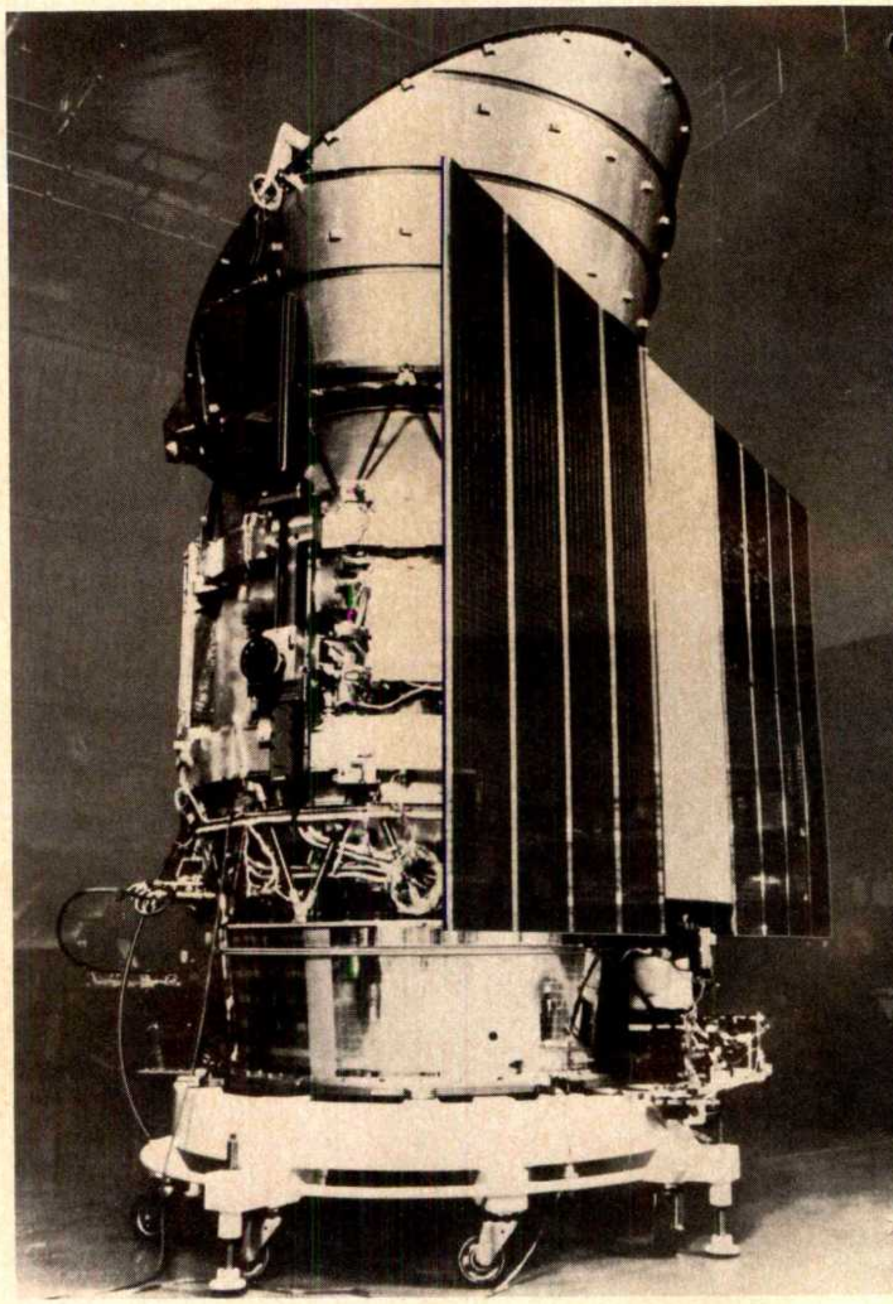
Space-age explorations of the heavens, using astronomical satellites to compile systematic catalogues of radio, infrared, ultraviolet and X-ray radiations, have fundamentally changed man's existing picture of the universe.

An example of such pioneering sky maps was the inventory of X-ray and ultraviolet cosmic sources made by the Astronomical Netherlands Satellite, ANS, a few years ago. The celestial map drawn on the basis of the measurements made is still being used for detailed observations.

The technological lead gained by ANS will soon be followed-up by placing another Dutch astronomical satellite, IRAS, into a 900 km orbit around the earth. Basically, the task of IRAS is to compile a sky map covering all infrared radiation between 8 and 120 μm, as well as to measure the intensity of radiation from the infrared sources.

Having made an important contribution to the success of the ANS space platform, Philips has again been chosen to provide technological assistance for this new, equally ambitious astronomical project. Responsibilities include IRAS's on-board computer, power supply, communication system and sub-systems integration and testing.

The requirements of satellite technology, both for in space and on the ground, represent a formidable challenge to the technical and organisational abilities of the participating industries. And, by integrating the experience, expertise and technological resources of its specialized divisions and member companies, Philips is continuing to meet this challenge – as the following examples illustrate.



PHILIPS

nature

Vol. 297 No. 5863 20 May 1982

Cover
Dispersal patterns of effluent from hot springs on the East Pacific Rise — see page 187.

Nature® (ISSN 0028-0836) is published weekly on Thursday, except the last week in December, by Macmillan Journals Ltd and includes the Directory of Biologicals (mailed in December) and Annual Index (mailed in February). Annual subscription for USA and Canada US $198.50 (for subscription prices elsewhere, see next page). Orders (with remittance) and change of address labels to: Macmillan Journals Ltd, Brunel Rd, Basingstoke RG21 2XS, UK. Second class postage paid at New York, NY10010 and additional mailing offices. US Postmaster send form 3579 to: Nature, 15 East 26 Street, New York, NY 10010.

MISCELLANY

GUIDE TO AUTHORS

Authors should be aware of the diversity of *Nature*'s readership and should strive to be as widely understood as possible.

Review articles should be accessible to the whole readership. Most are commissioned, but unsolicited reviews are welcome (in which case prior consultation with the office is desirable).

Scientific articles are research reports whose conclusions are of general interest or which represent substantial advances of understanding. The text should not exceed 3,000 words and six displayed items (figures plus tables). The article should include an abstract of about 50 words.

Letters to *Nature* are ordinarily 1,000 words long with no more than four displayed items. The first paragraph (not exceeding 150 words) should say what the letter is about, why the study it reports was undertaken and what the conclusions are.

Matters arising are brief comments (up to 500 words) on articles and letters recently published in *Nature*. The originator of a Matters Arising contribution should initially send his manuscript to the author of the original paper and both parties should, wherever possible, agree on what is to be submitted.

Manuscripts may be submitted either to London or Washington (decisions being made only in London). Manuscripts should be typed (double spacing) on one side of the paper only. Three copies are required, each accompanied by copies of lettered artwork. No title should exceed 80 characters in length. Reference lists, figure legends, etc. should be on separate sheets, all of which should be numbered. Abbreviations, symbols, units, etc. should be identified on one copy of the manuscript at their first appearance.

References should appear sequentially indicated by superscripts in the text and should be abbreviated according to the *World List of Scientific Periodicals*, fourth edition (Butterworth 1963–65). The first and last page numbers of each reference should be cited. References to books should clearly indicate the publisher and the date and place of publication. Unpublished articles should not be formally referred to unless accepted or submitted for publication, but may be mentioned in the text.

Each piece of artwork should be clearly marked with the author's name and the figure number. Original artwork should be unlettered. Suggestions for cover illustrations are welcome. Original artwork (and one copy of the manuscript) will be returned when manuscripts cannot be published.

Requests for permission to reproduce material from Nature *should be accompanied by a self-addressed (and, in the case of the UK and USA, stamped) envelope.*

Annual Subscription including Index (51 issues)

UK & Eire		£85
USA & Canada		US$198.50
Belgium	Airspeed only	BF7450
West Germany	Airspeed only	DM450
Netherlands	Airspeed only	G505
Switzerland	Airspeed only	SF395
Rest of Europe	Airspeed only	£100
Rest of World	Surface	£100
Rest of World*	Airmail	£150

**(not USA, Canada & Europe)*

Personal subscription rates are available in some countries to subscribers paying by personal cheque or credit card.

Details can be obtained from:
UK: *Nature* Promotion Department, Canada Road, Byfleet, Surrey KT14 7JL.
Telephone: Jonathan Earl: Byfleet (09323) 41459

USA: *Nature*, 15 East 26 Street, New York, NY 10010
Credit card orders only (in USA & Canada):
Call toll-free: (800) 824-7888 (Operator 246)
In California: (800) 852-7777 (Operator 246)

All other circulation enquiries concerning existing subscriptions please contact:
Nature Circulation Department, Macmillan Journals Ltd., Brunel Road, Basingstoke, Hants RG21 2XS, UK
Telephone: (0256) 29242 Telex: 858493

Back issues: (Post-paid) UK, £2.50; USA & Canada, US$6.00 (surface), US$9.00 (air); Rest of World, £3.00 (surface), £4.00 (air).

***Nature* Binders:**
Single Binders: UK, £4.00; Rest of World, $10.00
Set(s) of 3 Binders: UK, £10.00; Rest of World, 25.00

***Nature* First Issue Facsimile:**
UK, 75p; Rest of World (surface), $1.50, (air), £2.00

***Nature* Annual Indexes (1971-1980):**
Price (post-paid):
UK, £5.00 each; Rest of World, $10.00

***Nature* Directory of Biologicals:**
(Hardcover edition)
UK, £22.50; Rest of World, $45.00

***Nature* Wallcharts:**
How Britain Runs its Science; How the US Runs its Science; UK, £3.50; USA, $7.95 each

***Nature* in Microform:**
For Information:
UK: UMI, 18 Bedford Row, London WC2R 4EJ
USA: 300 North Zeeb Road, Ann Arbor, Mi 48106

Orders (with remittance) to:
Nature, Macmillan Journals Ltd., Brunel Road, Basingstoke, Hants RG21 2XS, UK
Tel: (0256) 29242 Telex: 858493

MACMILLAN JOURNALS

nature

Vol. 297 No. 5864 27 May 1982

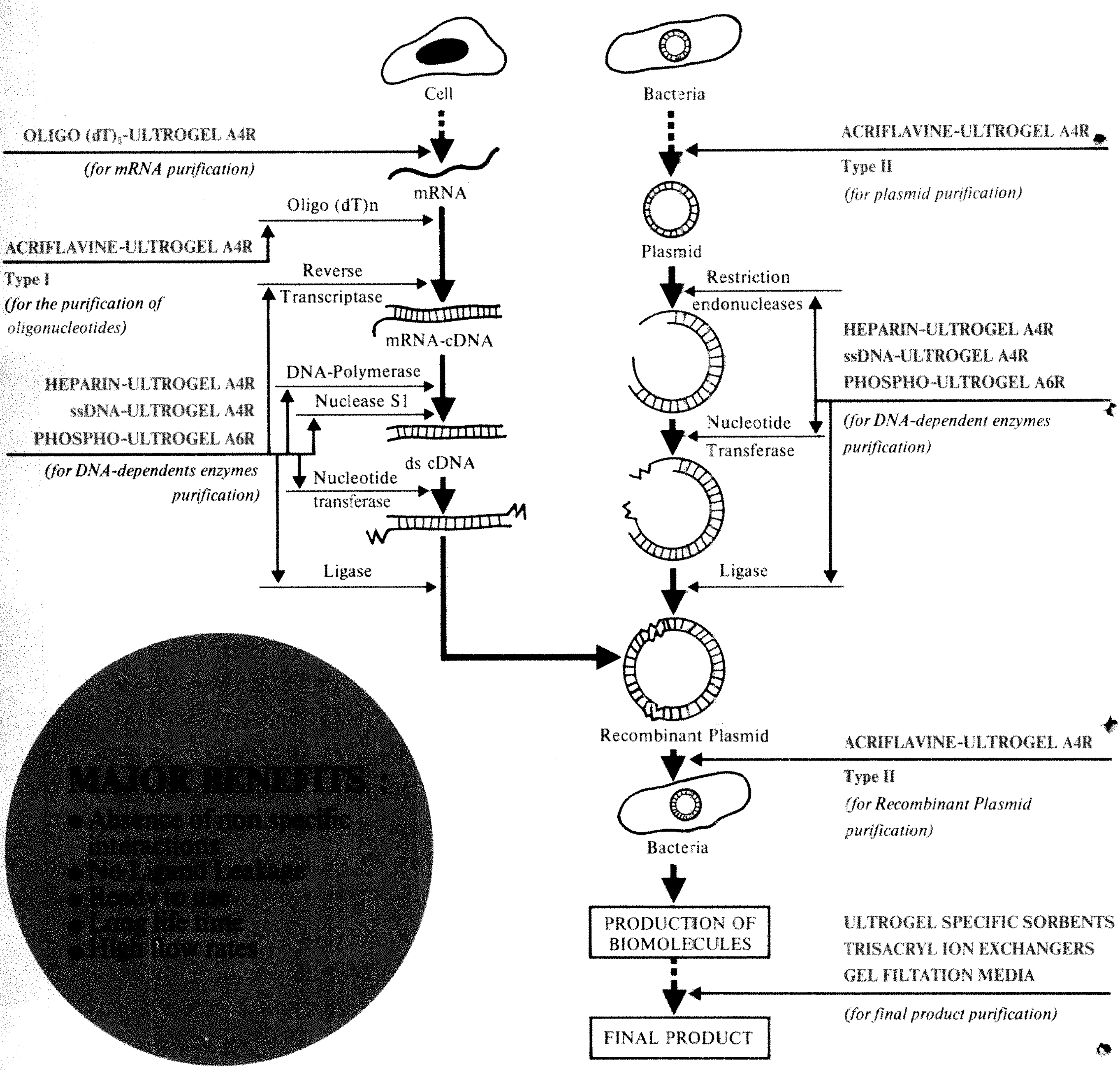
Cover
West Germany's flag and a view of the heavy ion accelerator at Darmstadt. See page **261**.

Nature® (ISSN 0028-0836) is published weekly on Thursday, except the last week in December, by Macmillan Journals Ltd and includes the Directory of Biologicals (mailed in December) and Annual Index (mailed in February). Annual subscription for USA and Canada US $198.50 (for subscription prices elsewhere, see next page). Orders (with remittance) and change of address labels to: Macmillan Journals Ltd, Brunel Rd, Basingstoke RG21 2XS, UK. Second class postage paid at New York, NY10010 and additional mailing offices. US Postmaster send form 3579 to: *Nature*, 15 East 26 Street, New York, NY 10010. ©1982 Macmillan Journals Ltd.

0028-0836/82/21000C-02$01.00

Contents continued overleaf

IBF affinity chromatography for molecular genetics



MAJOR BENEFITS :
- Absence of non specific interactions
- No Ligand Leakage
- Ready to use
- Long life time
- High flow rates

– « An IBF support for every chromatographic step in Molecular Genetic » –

Réactifs IBF

For France : **Pointet Girard** Département **Réactifs IBF**, 35, avenue Jean-Jaurès – 92390 Villeneuve-la-Garenne.

For other countries : please contact your local LKB representative.

Circle No.22 on Reader Service Card.

Annual Subscription including Index (51 issues)

UK & Eire		£85
USA & Canada		US$198.50
Belgium	Airspeed only	BF7450
West Germany	Airspeed only	DM450
Netherlands	Airspeed only	G505
Switzerland	Airspeed only	SF395
Rest of Europe	Airspeed only	£100
Rest of World	Surface	£100
Rest of World*	Airmail	£150

**(not USA, Canada & Europe)*

Personal subscription rates are available in some countries to subscribers paying by personal cheque or credit card.
Details can be obtained from:
UK: *Nature* Promotion Department, Canada Road, Byfleet, Surrey KT14 7JL.
Telephone: Jonathan Earl: Byfleet (09323) 41459
USA: *Nature*, 15 East 26 Street, New York, NY 10010
Credit card orders only (in USA & Canada):
Call toll-free: (800) 824-7888 (Operator 246)
In California: (800) 852-7777 (Operator 246)
All other circulation enquiries concerning existing subscriptions please contact:
Nature Circulation Department, Macmillan Journals Ltd., Brunel Road, Basingstoke, Hants RG21 2XS, UK
Telephone: (0256) 29242 Telex: 858493
Back issues: (Post-paid) UK, £2.50; USA & Canada, US$6.00 (surface), US$9.00 (air); Rest of World, £3.00 (surface), £4.00 (air).
***Nature* Binders:**
Single Binders: UK, £4.00; Rest of World, $10.00
Set(s) of 3 Binders: UK, £10.00; Rest of World, 25.00
***Nature* First Issue Facsimile:**
UK, 75p; Rest of World (surface), $1.50, (air), £2.00
***Nature* Annual Indexes (1971-1980):**
Price (post-paid):
UK, £5.00 each; Rest of World, $10.00
***Nature* Directory of Biologicals:**
(Hardcover edition)
UK, £22.50; Rest of World, $45.00
***Nature* Wallcharts:**
How Britain Runs its Science; How the US Runs its Science; UK, £3.50; USA, $7.95 each
***Nature* in Microform:**
For Information:
UK: UMI, 18 Bedford Row, London WC2R 4EJ
USA: 300 North Zeeb Road, Ann Arbor, Mi 48106
Orders (with remittance) to:
Nature, Macmillan Journals Ltd., Brunel Road, Basingstoke, Hants RG21 2XS, UK
Tel: (0256) 29242 Telex: 858493

MACMILLAN JOURNALS

MISCELLANY

GUIDE TO AUTHORS

Authors should be aware of the diversity of *Nature*'s readership and should strive to be as widely understood as possible.

Review articles should be accessible to the whole readership. Most are commissioned, but unsolicited reviews are welcome (in which case prior consultation with the office is desirable).

Scientific articles are research reports whose conclusions are of general interest or which represent substantial advances of understanding. The text should not exceed 3,000 words and six displayed items (figures plus tables). The article should include an abstract of about 50 words.

Letters to *Nature* are ordinarily 1,000 words long with no more than four displayed items. The first paragraph (not exceeding 150 words) should say what the letter is about, why the study it reports was undertaken and what the conclusions are.

Matters arising are brief comments (up to 500 words) on articles and letters recently published in *Nature*. The originator of a Matters Arising contribution should initially send his manuscript to the author of the original paper and both parties should, wherever possible, agree on what is to be submitted.

Manuscripts may be submitted either to London or Washington (decisions being made only in London). Manuscripts should be typed (double spacing) on one side of the paper only. Three copies are required, each accompanied by copies of lettered artwork. No title should exceed 80 characters in length. Reference lists, figure legends, etc. should be on separate sheets, all of which should be numbered. Abbreviations, symbols, units, etc. should be identified on one copy of the manuscript at their first appearance.

References should appear sequentially indicated by superscripts in the text and should be abbreviated according to the *World List of Scientific Periodicals*, fourth edition (Butterworth 1963–65). The first and last page numbers of each reference should be cited. References to books should clearly indicate the publisher and the date and place of publication. Unpublished articles should not be formally referred to unless accepted or submitted for publication, but may be mentioned in the text.

Each piece of artwork should be clearly marked with the author's name and the figure number. Original artwork should be unlettered. Suggestions for cover illustrations are welcome. Original artwork (and one copy of the manuscript) will be returned when manuscripts cannot be published.

Requests for permission to reproduce material from Nature *should be accompanied by a self-addressed (and, in the case of the UK and USA, stamped) envelope.*

nature

Vol. 297 No. 5865 3 June 1982

Cover
The 1981 eruption of the Le Piton de la Fournaise volcano is described on page 395

Nature® (ISSN 0028-0836) is published weekly on Thursday, except the last week in December, by Macmillan Journals Ltd and includes the Directory of Biologicals (mailed in December) and Annual Index (mailed in February). Annual subscription for USA and Canada US $198.50 (for subscription prices elsewhere, see next page). Orders (with remittance) and change of address labels to: Macmillan Journals Ltd, Brunel Rd, Basingstoke RG21 2XS, UK. Second class postage paid at New York, NY10010 and additional mailing offices. US Postmaster send form 3579 to: *Nature*, 15 East 26 Street, New York, NY 10010.

0028-0836/82/22000C-02$01.00

Contents continued overleaf

Annual Subscription including Index (51 issues)

UK & Eire		£85
USA & Canada		US$198.50
Belgium	Airspeed only	BF7450
West Germany	Airspeed only	DM450
Netherlands	Airspeed only	G505
Switzerland	Airspeed only	SF395
Rest of Europe	Airspeed only	£100
Rest of World	Surface	£100
Rest of World*	Airmail	£150

(not USA, Canada & Europe)

Personal subscription rates are available in some countries to subscribers paying by personal cheque or credit card.
Details can be obtained from:
UK: *Nature* Promotion Department, Canada Road, Byfleet, Surrey KT14 7JL. Telephone: Jonathan Earl: Byfleet (09323) 41459
USA: *Nature*, 15 East 26 Street, New York, NY 10010
Credit card orders only (in USA & Canada):
Call toll-free: (800) 824-7888 (Operator 246)
In California: (800) 852-7777 (Operator 246)
All other circulation enquiries concerning existing subscriptions please contact:
Nature Circulation Department, Macmillan Journals Ltd., Brunel Road, Basingstoke, Hants RG21 2XS, UK
Telephone: (0256) 29242 Telex: 858493
Back issues: (Post-paid) UK, £2.50; USA & Canada, US$6.00 (surface), US$9.00 (air); Rest of World, £3.00 (surface), £4.00 (air).
***Nature* Binders:**
Single Binders: UK, £4.00; Rest of World, $10.00
Set(s) of 3 Binders: UK, £10.00; Rest of World, 25.00
***Nature* First Issue Facsimile:**
UK, 75p; Rest of World (surface), $1.50, (air), £2.00
***Nature* Annual Indexes (1971-1980):**
Price (post-paid):
UK, £5.00 each; Rest of World, $10.00
***Nature* Directory of Biologicals:**
(Hardcover edition)
UK, £22.50; Rest of World, $45.00
***Nature* Wallcharts:**
How Britain Runs its Science; How the US Runs its Science; UK, £3.50; USA, $7.95 each
***Nature* in Microform:**
For Information:
UK: UMI, 18 Bedford Row, London WC2R 4EJ
USA: 300 North Zeeb Road, Ann Arbor, Mi 48106
Orders (with remittance) to:
Nature, Macmillan Journals Ltd., Brunel Road, Basingstoke, Hants RG21 2XS, UK
Tel: (0256) 29242 Telex: 858493

MACMILLAN JOURNALS

GUIDE TO AUTHORS

Authors should be aware of the diversity of *Nature*'s readership and should strive to be as widely understood as possible.

Review articles should be accessible to the whole readership. Most are commissioned, but unsolicited reviews are welcome (in which case prior consultation with the office is desirable).

Scientific articles are research reports whose conclusions are of general interest or which represent substantial advances of understanding. The text should not exceed 3,000 words and six displayed items (figures plus tables). The article should include an abstract of about 50 words.

Letters to *Nature* are ordinarily 1,000 words long with no more than four displayed items. The first paragraph (not exceeding 150 words) should say what the letter is about, why the study it reports was undertaken and what the conclusions are.

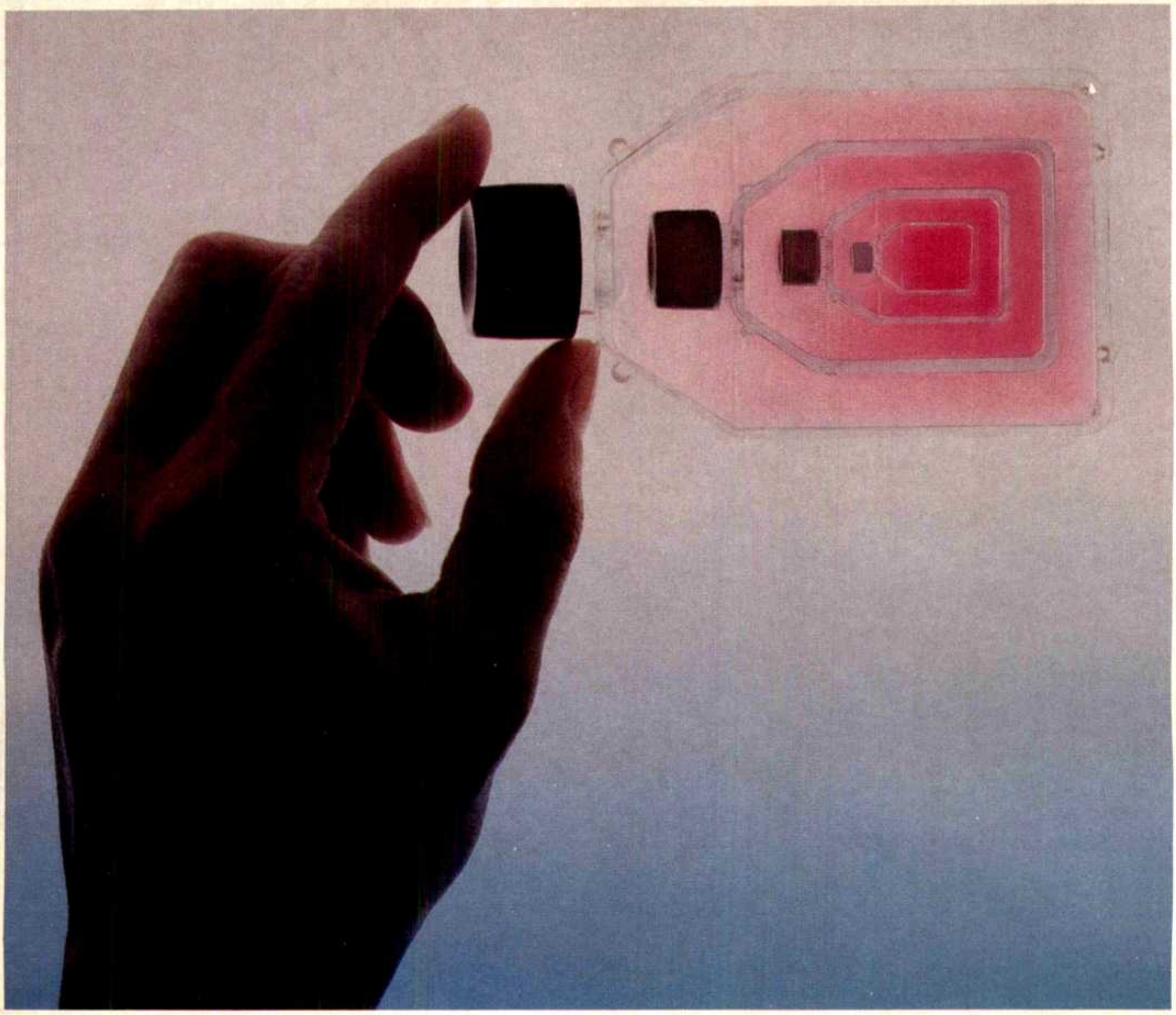
Matters arising are brief comments (up to 500 words) on articles and letters recently published in *Nature*. The originator of a Matters Arising contribution should initially send his manuscript to the author of the original paper and both parties should, wherever possible, agree on what is to be submitted.

Manuscripts may be submitted either to London or Washington (decisions being made only in London). Manuscripts should be typed (double spacing) on one side of the paper only. Three copies are required, each accompanied by copies of lettered artwork. No title should exceed 80 characters in length. Reference lists, figure legends, etc. should be on separate sheets, all of which should be numbered. Abbreviations, symbols, units, etc. should be identified on one copy of the manuscript at their first appearance.

References should appear sequentially indicated by superscripts in the text and should be abbreviated according to the *World List of Scientific Periodicals*, fourth edition (Butterworth 1963–65). The first and last page numbers of each reference should be cited. References to books should clearly indicate the publisher and the date and place of publication. Unpublished articles should not be formally referred to unless accepted or submitted for publication, but may be mentioned in the text.

Each piece of artwork should be clearly marked with the author's name and the figure number. Original artwork should be unlettered. Suggestions for cover illustrations are welcome. Original artwork (and one copy of the manuscript) will be returned when manuscripts cannot be published.

Requests for permission to reproduce material from Nature *should be accompanied by a self-addressed (and, in the case of the UK and USA, stamped) envelope.*



nature

Vol. 297 No. 5866 10 June 1982

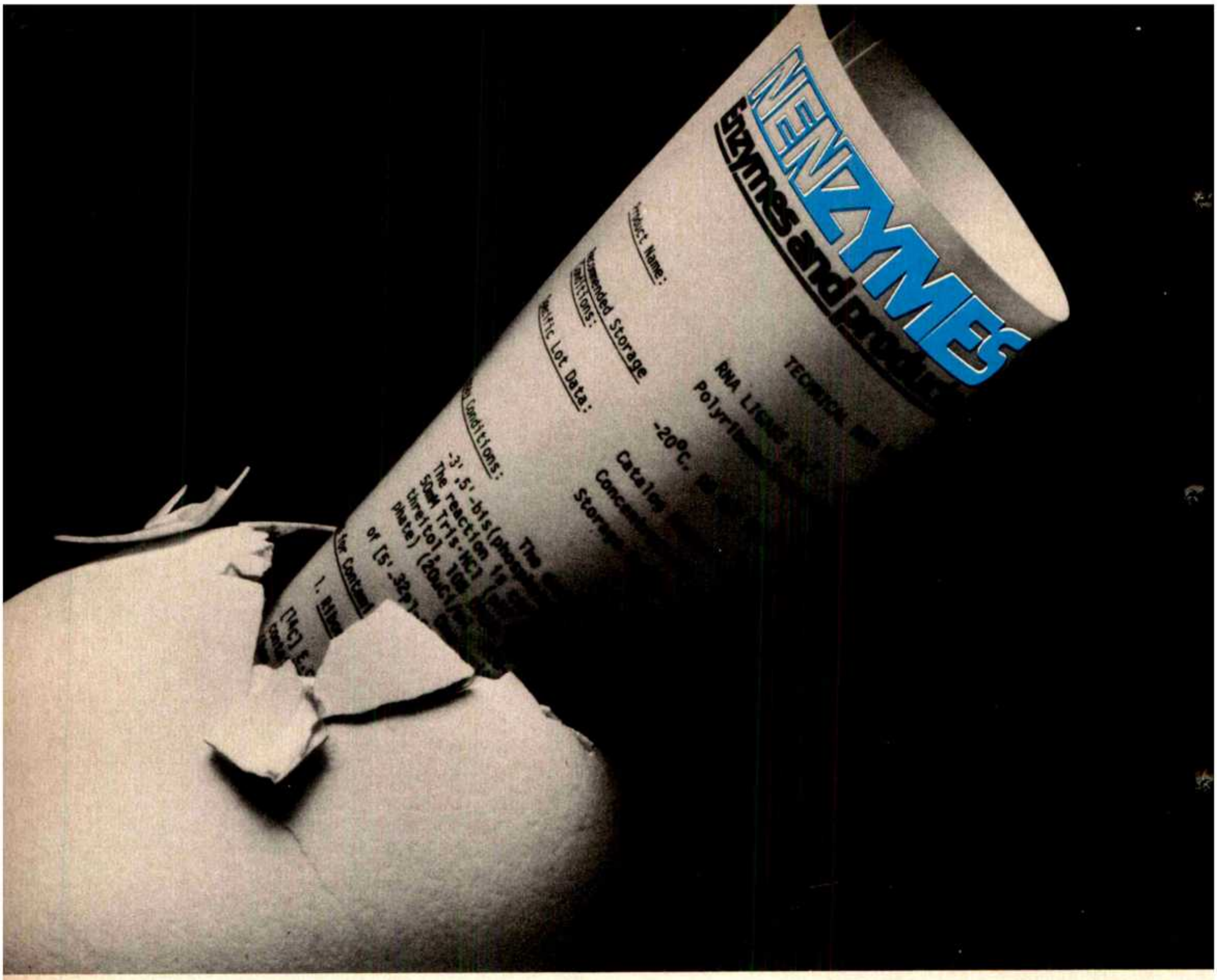
Cover

A thunderstorm on the Great Plains of North America. The use of Doppler weather radar to study storms and lightning is reviewed on page 461

Nature® (ISSN 0028-0836) is published weekly on Thursday, except the last week in December, by Macmillan Journals Ltd and includes the Directory of Biologicals (mailed in December) and Annual Index (mailed in February). Annual subscription for USA and Canada US $198.50 (for subscription prices elsewhere, see next page). Orders (with remittance) and change of address labels to: Macmillan Journals Ltd, Brunel Rd, Basingstoke RG21 2XS, UK. Second class postage paid at New York, NY10010 and additional mailing offices. US Postmaster send form 3579 to: *Nature*, 15 East 26 Street, New York, NY 10010. ©1982 Macmillan Journals Ltd.

0028-0836/82/23000C-02$01.00

Contents continued overleaf



Breakthrough in biotechnology

Just when research needed it, NENZYMES™ provided it. Now more and more people doing DNA-RNA studies are using these enzymes — because they are ultra pure. NENZYMES eliminate one variable from the many in molecular biology.

NENZYMES are tested for biological purity in two stages: First, in numerous specific assays of each produced lot (which are recorded for customer inspection); and again when used as integral components in our pre-packaged systems for DNA/RNA research. They have to be trouble-free before they go to you.

RNA Ligase is the latest addition to our line of DNA/RNA modifying enzymes. We also offer restriction enzymes as well as selected nucleic acids — all of the highest quality. Call us toll free for complete information on any of these products.

New England Nuclear, 549 Albany Street, Boston, MA 02118
Call toll free: 800-225-1572, Telex: 94-0996
Mass and Internat'l: 617-482-9595
Europe: NEN Chemicals GmbH, D-6072, W. Germany
Postfach 401240, Tel. (06103) 803-0, Telex 4-17993 NEN D
NEN Canada: 2453 46th Avenue, Lachine, Que H8T 3C9
Tel. 514-636-4971, Telex 05-821808

Circle No.08 on Reader Service Card.

New England Nuclear®
a Du Pont company

Annual Subscription including Index (51 issues)

UK & Eire		£85
USA & Canada		US$198.50
Belgium	Airspeed only	BF7450
West Germany	Airspeed only	DM450
Netherlands	Airspeed only	G505
Switzerland	Airspeed only	SF395
Rest of Europe	Airspeed only	£100
Rest of World	Surface	£100
Rest of World*	Airmail	£150

**(not USA, Canada & Europe)*

Personal subscription rates are available in some countries to subscribers paying by personal cheque or credit card.

Details can be obtained from:
UK: *Nature* Promotion Department, Canada Road, Byfleet, Surrey KT14 7JL. Telephone: Jonathan Earl: Byfleet (09323) 41459
USA: *Nature*, 15 East 26 Street, New York, NY 10010
Credit card orders only (in USA & Canada): Call toll-free: (800) 824-7888 (Operator 246)
In California: (800) 852-7777 (Operator 246)

All other circulation enquiries concerning existing subscriptions please contact:
Nature Circulation Department, Macmillan Journals Ltd., Brunel Road, Basingstoke, Hants RG21 2XS, UK
Telephone: (0256) 29242 Telex: 858493

Back issues: (Post-paid) UK, £2.50; USA & Canada, US$6.00 (surface), US$9.00 (air); Rest of World, £3.00 (surface), £4.00 (air).

***Nature* Binders:**
Single Binders: UK, £4.00; Rest of World, $10.00
Set(s) of 3 Binders: UK, £10.00; Rest of World, 25.00

***Nature* First Issue Facsimile:**
UK, 75p; Rest of World (surface), $1.50, (air), £2.00

***Nature* Annual Indexes (1971-1980):**
Price (post-paid):
UK, £5.00 each; Rest of World, $10.00

***Nature* Directory of Biologicals:**
(Hardcover edition)
UK, £22.50; Rest of World, $45.00

***Nature* Wallcharts:**
How Britain Runs its Science; How the US Runs its Science; UK, £3.50; USA, $7.95 each

***Nature* in Microform:**
For Information:
UK: UMI, 18 Bedford Row, London WC2R 4EJ
USA: 300 North Zeeb Road, Ann Arbor, Mi 48106

Orders (with remittance) to:
Nature, Macmillan Journals Ltd., Brunel Road, Basingstoke, Hants RG21 2XS, UK
Tel: (0256) 29242 Telex: 858493

MACMILLAN JOURNALS

MISCELLANY

GUIDE TO AUTHORS

Authors should be aware of the diversity of *Nature*'s readership and should strive to be as widely understood as possible.

Review articles should be accessible to the whole readership. Most are commissioned, but unsolicited reviews are welcome (in which case prior consultation with the office is desirable).

Scientific articles are research reports whose conclusions are of general interest or which represent substantial advances of understanding. The text should not exceed 3,000 words and six displayed items (figures plus tables). The article should include an abstract of about 50 words.

Letters to *Nature* are ordinarily 1,000 words long with no more than four displayed items. The first paragraph (not exceeding 150 words) should say what the letter is about, why the study it reports was undertaken and what the conclusions are.

Matters arising are brief comments (up to 500 words) on articles and letters recently published in *Nature*. The originator of a Matters Arising contribution should initially send his manuscript to the author of the original paper and both parties should, wherever possible, agree on what is to be submitted.

Manuscripts may be submitted either to London or Washington (decisions being made only in London). Manuscripts should be typed (double spacing) on one side of the paper only. Three copies are required, each accompanied by copies of lettered artwork. No title should exceed 80 characters in length. Reference lists, figure legends, etc. should be on separate sheets, all of which should be numbered. Abbreviations, symbols, units, etc. should be identified on one copy of the manuscript at their first appearance.

References should appear sequentially indicated by superscripts in the text and should be abbreviated according to the *World List of Scientific Periodicals*, fourth edition (Butterworth 1963–65). The first and last page numbers of each reference should be cited. References to books should clearly indicate the publisher and the date and place of publication. Unpublished articles should not be formally referred to unless accepted or submitted for publication, but may be mentioned in the text.

Each piece of artwork should be clearly marked with the author's name and the figure number. Original artwork should be unlettered. Suggestions for cover illustrations are welcome. Original artwork (and one copy of the manuscript) will be returned when manuscripts cannot be published.

Requests for permission to reproduce material from Nature *should be accompanied by a self-addressed (and, in the case of the UK and USA, stamped) envelope.*

Nature in Poland

Readers wishing to subscribe to a weekly copy of *Nature* on behalf of a colleague in Poland (where hard currency is scarce) are directed to the special offer in last week's *Nature*.

nature

Vol. 297 No. 5867 17 June 1982

Cover

Thin section of pumice from Great Sitkin Island in the Aleutian Arc showing evidence of intimate mixing between andesitic magma and dacitic magma. See page 554

Nature® (ISSN 0028-0836) is published weekly on Thursday, except the last week in December, by Macmillan Journals Ltd and includes the Directory of Biologicals (mailed in December) and Annual Index (mailed in February). Annual subscription for USA and Canada US $198.50 (for subscription prices elsewhere, see next page). Orders (with remittance) and change of address labels to: Macmillan Journals Ltd, Brunel Rd, Basingstoke RG21 2XS, UK. Second class postage paid at New York, NY10010 and additional mailing offices. US Postmaster send form 3579 to: *Nature*, 15 East 26 Street, New York, NY 10010.

0028-0836/82/24000C-02$01.00

Contents continued overleaf

The New Way of Separating

- *plasmids*
- *DNA fragments*
- *viruses*
- *sub-cellular particles*
- *large proteoglycans and polysaccharides*

– by gel filtration

Sephacryl® S-1000 Superfine is a unique new gel which offers a rapid, inexpensive and mild method for the separation of large molecules by gel filtration. A combination of extremely high porosity and good rigidity allows you to separate large molecules and particles up to 300–400 nm in diameter.

Put Sephacryl S-1000 in your lab and reduce your time spent in the centrifuge room! Technical information and detailed instructions are available upon request from Pharmacia Fine Chemicals.

Pharmacia Fine Chemicals
Division of Pharmacia Inc.
800 Centennial Avenue
Piscataway, N.J. 08854

Pharmacia Fine Chemicals

Annual Subscription including Index (51 issues)

UK & Eire		£85
USA & Canada		US$198.50
Belgium	Airspeed only	BF7450
West Germany	Airspeed only	DM450
Netherlands	Airspeed only	G505
Switzerland	Airspeed only	SF395
Rest of Europe	Airspeed only	£100
Rest of World	Surface	£100
Rest of World*	Airmail	£150

(not USA, Canada & Europe)

Personal subscription rates are available in some countries to subscribers paying by personal cheque or credit card.
Details can be obtained from:
UK: *Nature* Promotion Department, Canada Road, Byfleet, Surrey KT14 7JL.
Telephone: Jonathan Earl: Byfleet (09323) 41459
USA: *Nature*, 15 East 26 Street, New York, NY 10010
Credit card orders only (in USA & Canada):
Call toll-free: (800) 824-7888 (Operator 246)
In California: (800) 852-7777 (Operator 246)
All other circulation enquiries concerning existing subscriptions please contact:
Nature Circulation Department, Macmillan Journals Ltd., Brunel Road, Basingstoke, Hants RG21 2XS, UK
Telephone: (0256) 29242 Telex: 858493
Back issues: (Post-paid) UK, £2.50; USA & Canada, US$6.00 (surface), US$9.00 (air); Rest of World, £3.00 (surface), £4.00 (air).
***Nature* Binders:**
Single Binders: UK, £4.00; Rest of World, $10.00
Set(s) of 3 Binders: UK, £10.00; Rest of World, 25.00
***Nature* First Issue Facsimile:**
UK, 75p; Rest of World (surface), $1.50, (air), £2.00
***Nature* Annual Indexes (1971-1980):**
Price (post-paid):
UK, £5.00 each; Rest of World, $10.00
***Nature* Directory of Biologicals:**
(Hardcover edition)
UK, £22.50; Rest of World, $45.00
***Nature* Wallcharts:**
How Britain Runs its Science; How the US Runs its Science; UK, £3.50; USA, $7.95 each
***Nature* in Microform:**
For Information:
UK: UMI, 18 Bedford Row, London WC2R 4EJ
USA: 300 North Zeeb Road, Ann Arbor, Mi 48106
Orders (with remittance) to:
Nature, Macmillan Journals Ltd., Brunel Road, Basingstoke, Hants RG21 2XS, UK
Tel: (0256) 29242 Telex: 858493

MACMILLAN JOURNALS

GUIDE TO AUTHORS

Authors should be aware of the diversity of *Nature*'s readership and should strive to be as widely understood as possible.

Review articles should be accessible to the whole readership. Most are commissioned, but unsolicited reviews are welcome (in which case prior consultation with the office is desirable).

Scientific articles are research reports whose conclusions are of general interest or which represent substantial advances of understanding. The text should not exceed 3,000 words and six displayed items (figures plus tables). The article should include an abstract of about 50 words.

Letters to *Nature* are ordinarily 1,000 words long with no more than four displayed items. The first paragraph (not exceeding 150 words) should say what the letter is about, why the study it reports was undertaken and what the conclusions are.

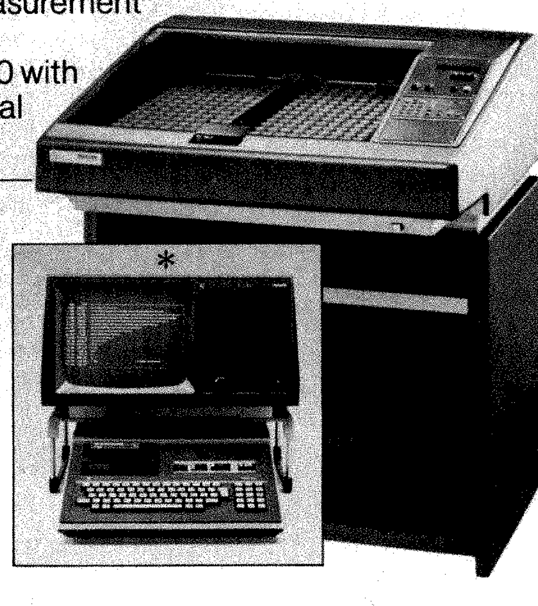
Matters arising are brief comments (up to 500 words) on articles and letters recently published in *Nature*. The originator of a Matters Arising contribution should initially send his manuscript to the author of the original paper and both parties should, wherever possible, agree on what is to be submitted.

Manuscripts may be submitted either to London or Washington (decisions being made only in London). Manuscripts should be typed (double spacing) on one side of the paper only. Three copies are required, each accompanied by copies of lettered artwork. No title should exceed 80 characters in length. Reference lists, figure legends, etc. should be on separate sheets, all of which should be numbered. Abbreviations, symbols, units, etc. should be identified on one copy of the manuscript at their first appearance.

References should appear sequentially indicated by superscripts in the text and should be abbreviated according to the *World List of Scientific Periodicals*, fourth edition (Butterworth 1963–65). The first and last page numbers of each reference should be cited. References to books should clearly indicate the publisher and the date and place of publication. Unpublished articles should not be formally referred to unless accepted or submitted for publication, but may be mentioned in the text.

Each piece of artwork should be clearly marked with the author's name and the figure number. Original artwork should be unlettered. Suggestions for cover illustrations are welcome. Original artwork (and one copy of the manuscript) will be returned when manuscripts cannot be published.

Requests for permission to reproduce material from Nature *should be accompanied by a self-addressed (and, in the case of the UK and USA, stamped) envelope.*



nature

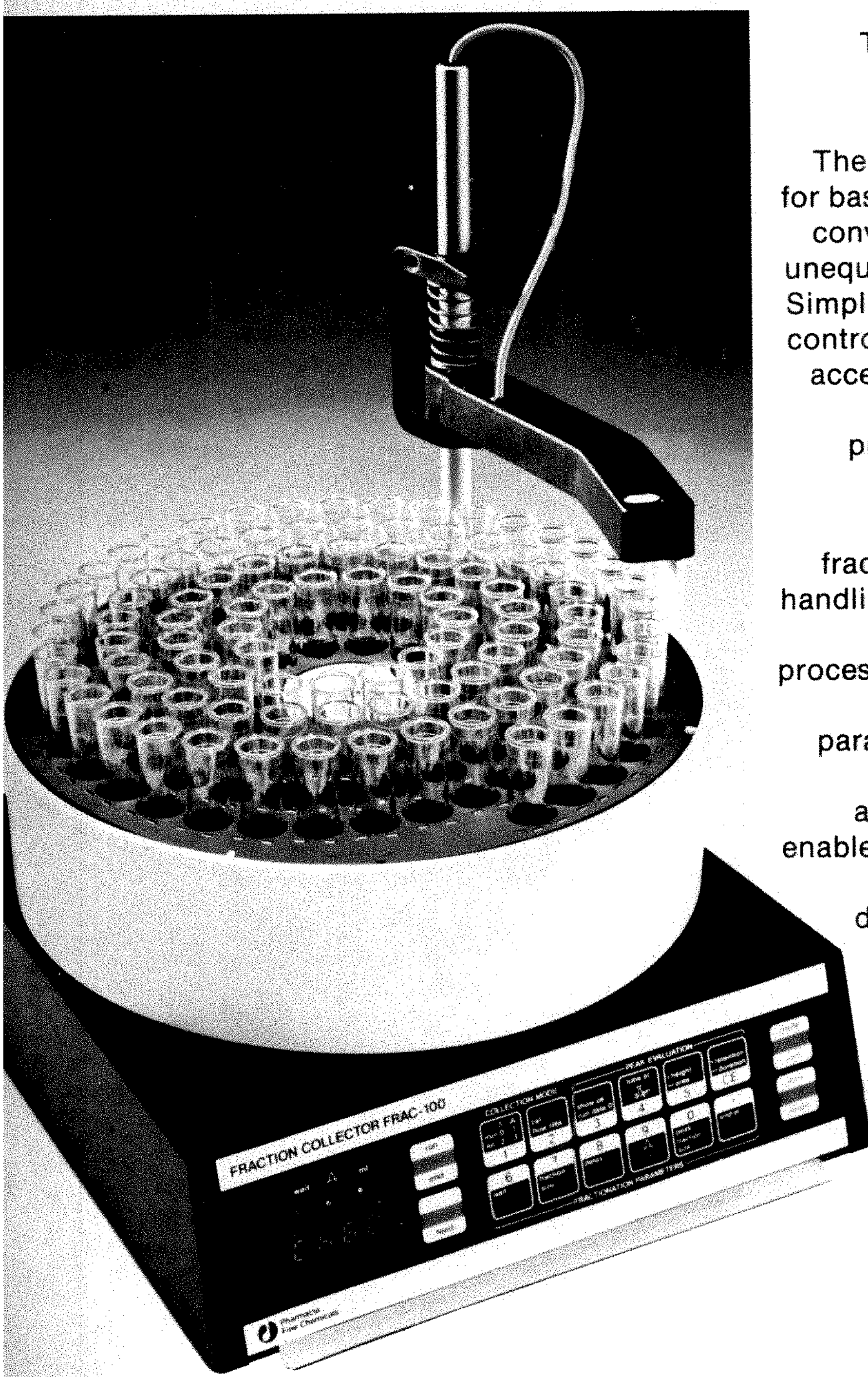
Vol. 297 No. 5868 24 June 1982

Nature® (ISSN 0028-0836) is published weekly on Thursday, except the last week in December, by Macmillan Journals Ltd and includes the Directory of Biologicals (mailed in December) and Annual Index (mailed in February). Annual subscription for USA and Canada US $198.50 (for subscription prices elsewhere, see next page). Orders (with remittance) and change of address labels to: Macmillan Journals Ltd, Brunel Rd, Basingstoke RG21 2XS, UK. Second class postage paid at New York, NY10010 and additional mailing offices. US Postmaster send form 3579 to: *Nature*, 15 East 26 Street, New York, NY 10010.

0028-0836/82/25000C-02$01.00

Contents continued overleaf

The FRAC-100 Fraction Collector



The FRAC-100 does what any other small fraction collector does - collects fractions, PLUS .....

The FRAC-100 will fulfill all your needs for basic fractionation requirements. As a conventional collector the FRAC-100 is unequaled in performance and reliability. Simplicity in design (3 moving parts) and control (flexible fractionation parameters accessed by a multifunction key board) accomplish fractionation routines previously attainable only with large, expensive instruments.

The FRAC-100, as a sophisticated fraction collector presents unique data handling and integration capabilities. The input of a 0-10mV detector signal is processed by the FRAC-100 to report as a digital display such separation parameters as peak height, peak width, peak retention and peak area. The addition of a PSV-100 solenoid valve enables selective fractionation of only the peak material as judged by a slope detector defined by a programmable threshold value.

If what you're looking for is a small, simple, inexpensive fraction collector - the FRAC-100 is your best choice!

If what you're looking for is a small sophisticated collector with power and abilities far beyond anything presently available - the FRAC-100 is your <u>only</u> choice!

Pharmacia Fine Chemicals
Division of Pharmacia Inc.
800 Centennial Avenue
Piscataway, N.J. 08854

Pharmacia Fine Chemicals

Circle No.42 on Reader Service Card.

Annual Subscription including Index (51 issues)

UK & Eire		£85
USA & Canada		US$198.50
Belgium	Airspeed only	BF7450
West Germany	Airspeed only	DM450
Netherlands	Airspeed only	G505
Switzerland	Airspeed only	SF395
Rest of Europe	Airspeed only	£100
Rest of World	Surface	£100
Rest of World*	Airmail	£150

**(not USA, Canada & Europe)*

Personal subscription rates are available in some countries to subscribers paying by personal cheque or credit card.
Details can be obtained from:
UK: *Nature* Promotion Department, Canada Road, Byfleet, Surrey KT14 7JL. Telephone: Jonathan Earl: Byfleet (09323) 41459
USA: *Nature*, 15 East 26 Street, New York, NY 10010
Credit card orders only (in USA & Canada):
Call toll-free: (800) 824-7888 (Operator 246)
In California: (800) 852-7777 (Operator 246)
All other circulation enquiries concerning existing subscriptions please contact:
Nature Circulation Department, Macmillan Journals Ltd., Brunel Road, Basingstoke, Hants RG21 2XS, UK
Telephone: (0256) 29242 Telex: 858493
Back issues: (Post-paid) UK, £2.50; USA & Canada, US$6.00 (surface), US$9.00 (air); Rest of World, £3.00 (surface), £4.00 (air).
***Nature* Binders:**
Single Binders: UK, £4.00; Rest of World, $10.00
Set(s) of 3 Binders: UK, £10.00; Rest of World, 25.00
***Nature* First Issue Facsimile:**
UK, 75p; Rest of World (surface), $1.50, (air), £2.00
***Nature* Annual Indexes (1971-1980):**
Price (post-paid):
UK, £5.00 each; Rest of World, $10.00
***Nature* Directory of Biologicals:**
(Hardcover edition)
UK, £22.50; Rest of World, $45.00
***Nature* Wallcharts:**
How Britain Runs its Science; How the US Runs its Science; UK, £3.50; USA, $7.95 each
***Nature* in Microform:**
For Information:
UK: UMI, 18 Bedford Row, London WC2R 4EJ
USA: 300 North Zeeb Road, Ann Arbor, Mi 48106
Orders (with remittance) to:
Nature, Macmillan Journals Ltd., Brunel Road, Basingstoke, Hants RG21 2XS, UK
Tel: (0256) 29242 Telex: 858493

MACMILLAN JOURNALS

GUIDE TO AUTHORS

Authors should be aware of the diversity of *Nature*'s readership and should strive to be as widely understood as possible.

Review articles should be accessible to the whole readership. Most are commissioned, but unsolicited reviews are welcome (in which case prior consultation with the office is desirable).

Scientific articles are research reports whose conclusions are of general interest or which represent substantial advances of understanding. The text should not exceed 3,000 words and six displayed items (figures plus tables). The article should include an abstract of about 50 words.

Letters to *Nature* are ordinarily 1,000 words long with no more than four displayed items. The first paragraph (not exceeding 150 words) should say what the letter is about, why the study it reports was undertaken and what the conclusions are.

Matters arising are brief comments (up to 500 words) on articles and letters recently published in *Nature*. The originator of a Matters Arising contribution should initially send his manuscript to the author of the original paper and both parties should, wherever possible, agree on what is to be submitted.

Manuscripts may be submitted either to London or Washington (decisions being made only in London). Manuscripts should be typed (double spacing) on one side of the paper only. Three copies are required, each accompanied by copies of lettered artwork. No title should exceed 80 characters in length. Reference lists, figure legends, etc. should be on separate sheets, all of which should be numbered. Abbreviations, symbols, units, etc. should be identified on one copy of the manuscript at their first appearance.

References should appear sequentially indicated by superscripts in the text and should be abbreviated according to the *World List of Scientific Periodicals*, fourth edition (Butterworth 1963–65). The first and last page numbers of each reference should be cited. References to books should clearly indicate the publisher and the date and place of publication. Unpublished articles should not be formally referred to unless accepted or submitted for publication, but may be mentioned in the text.

Each piece of artwork should be clearly marked with the author's name and the figure number. Original artwork should be unlettered. Suggestions for cover illustrations are welcome. Original artwork (and one copy of the manuscript) will be returned when manuscripts cannot be published.

Requests for permission to reproduce material from Nature *should be accompanied by a self-addressed (and, in the case of the UK and USA, stamped) envelope.*

nature

6 May 1982

How not to teach engineers

Changing the name of a research council will not help British engineering out of the rut. But what will?

British engineering education is once more in turmoil In spite of the appearance of the Finniston committee's report on engineering education just over two years ago, which went a long way to meeting previous discontents, and a more recent decision by the Science Research Council to add engineering to its formal title, the engineering professors from British universities, long known as a body of turbulent men, are still pressing for a separate engineering research council just for them The argument that inspires this movement has little substance and indeed is illogical if taken seriously What the engineering professors are doing is to combine elements from the Finniston report and their own practice which are consistent with each other but inconsistent with the needs of potential engineers, with the needs of British industry as they should be and indeed with the needs of education

The Science Research Council's decision to call itself the Science and Engineering Research Council seemed last year to be strictly cosmetic In reality, the Science Research Council had for many years been trying with conspicuous success to increase the proportion of the funds spent on engineering research in British universities The fact that over the years it has found this task very difficult is as much an indication of the receptiveness of the market — potential researchers in universities — to the availability of research funds as a sign that engineering is included in comparison with science and other fields of enquiry For the principle is that research is almost inevitably a less important part of British engineering education, at least as at present constituted, than of other scientific fields of enquiry (medicine perhaps excepted) For one thing (as the Finniston committee noted) many of those who enter engineering courses do so because they wish to join some part of the labour force at an early stage — wish to "get something done" Second, engineering research frequently entails use of large and expensive experimental rigs which, at least as British universities are at present organized, hard to come by Third, it is by no means clear whether research in a formal academic sense contributes to the creativity which the engineering professors, like the Finniston committee, consider to be the element in British engineering now most in need of being fostered In short, the case for regarding research as an essential part of engineering education rests only on the proper assumption that people who teach and do nothing else are probably poor teachers Yet, underpaid as they are, many British engineering academics prefer to spend what spare energy they have, not on the preparation of formal research reports, but on activities that may in the long run be just as useful to them as teachers — consulting for companies, working on real projects or perhaps even writing books that simplify scraps of information already known

So why is the cry that more funds should somehow be channelled into academic engineering research so insistent? The chief answer seems to be that the engineering professors and their followers have noted how, in other fields of science, people's prestige or even promotion depends on their formal completion of research projects of various kinds Finniston was persuaded of the importance of this model, suggesting that engineering would not acquire the status he rightly wished to give it until most engineering academics were able to brandish *curricula vitae* stuffed with research papers published in some formal body of literature The argument may be fair If this were indeed what most engineering academics could boast of, they would no doubt walk taller when mixing with their scientific colleagues in universities But it does not follow that the condition of engineering education would thereby be improved Indeed the Finniston report itself drew attention to the importance that engineering academics should be people who, first of all, were able to win the respect of their industrial rather than academic counterparts For, the argument goes, only then will their students respect them and listen to what they have to say

The same depressing conclusion is not equally applicable elsewhere, in many of the countries of Western Europe and the United States for example The reason for that difference, however, lies in one feature of the British engineering curriculum which needs more attention than it was given even by the Finniston committee Starting from its assumption about the motives persuading undergraduates to enter engineering courses in British universities, the Finniston committee argued that people wishing to work as engineers should be given plenty of opportunity to do so quickly In other words, the more a student was persuaded to acquire a good understanding of physical (and these days biological) science, the more likely he was to lose his sense that engineering is something different The result of that argument, accepted by the Finniston committee and most of those enthusiastic about its conclusions, is that even those engineering courses at British universities which have been "enriched" differ very little from the older courses to which have been grafted elements of the curriculum concerned with law, industrial relations and production engineering (whatever that may be) This implies that even if, as is likely, the next few years see a spread of these enriched courses throughout the British university system, engineers leaving the universities may still be ignorant of the rudiments of the sciences to which their skills apply Is this a recipe for providing the kinds of engineers likely to be at the forefront of industrial innovation? Or is it a device for making sure that even if the other obstacles to change in British industry were removed, engineers would still feel trapped by the past?

The Finniston report, rich as it was in comparisons between the economics of Britain and its industrial competitors, seems not fully to have heeded the implications of the curricular comparisons that it also included Elsewhere, especially in Western Europe, engineering education is indeed an intimate amalgam of sciences and its application To be sure courses are, by British standards, long — five rather than three or four years are commonplace The graduates from these more rigorous engineering schools appear to make a more direct contribution to the innovation of products and processes The Japanese example is perhaps even more striking — university engineers tend to be well trained in basic sciences and the companies which eventually employ them shoulder much of the task of helping them understand how their knowledge can be used Why should not the British educational system follow one or other of these paths? It

US contributors please note

From now on, authors in the United States and Canada should send their manuscripts either direct to London or to the Washington office, whose new address in *Nature,* 991 National Press Building, Washington, DC 20045, manuscripts should no longer be sent to *Nature*'s New York office

0028-0836/82/180001-02$01 00

would be wrong to blame the engineering professors, who appear willing to change, although they contribute to the suspicion growing between science and engineering in Britain Nor is the government to blame, for it seems fully aware that British engineering should be more innovative and more flexible The real difficulty is that for industry survival must now depend on making saleable products for the first time, and British industry does not fully understand what skilled engineers could accomplish

The Finniston report itself provided plenty of evidence to show that British engineers once at work have ahead of them a poor prospect for a career A few years after formal graduation, and once the two-year period required to become a chartered engineer is completed, the salaries of those still working as engineers for British companies appear to level off While the statistics are notoriously untrustworthy — many who began as engineers may quickly become salesmen or administrators — industrial companies appear to set little store by the work required of professional and experienced engineers But especially now, when the new technologies in microelectronics and telecommunications have thrown up almost an embarrassment of devices that might be manufactured and sold profitably, and when one of the most serious problems facing British companies and their competitors is to decide which of these might be made cheaply and sold profitably, there is a greater than ever need of people with a proper engineering education who might help chart a reliable course for the future What this in turn implies is that British industry must look for a system of higher education in engineering that will produce large numbers of qualified and creative people who will more often be employed in research and development than in the running of plants, production engineering or whatever and, if the engineering professors would reflect on that ideal state of affairs, they would surely recognize that then the status of engineers in Britain would truly be enhanced Contributing to the health of the economy would automatically win the respect that could not be acquired by a list of research papers an arm long It might then even be the case that engineering teachers, themselves transformed by industry's expectations of engineering qualifications, would more eagerly take up the funds that the Science and Engineering Research Council is apparently eager to seize back But another research council, exclusively concerned with engineering research would be at best a red herring

Economic research?

Can Congress realize its adage that investment in basic research will aid economic growth?

The US Congress is now wrestling with a problem common to all the industrial democracies how to get the most economic benefit out of the nation's investment in research and development Because the US economy is so decentralized, Congress is attacking the problem with the only central mechanism under its control, the federal research budget worked out each year between the current Administration and Congress Last December the US Congress, on a sweeping 90 to 0 vote, decided that one way to do this was to set aside one per cent of federal research and development funds (excluding those going to in-house research) for small independently-owned companies employing 500 people or less These firms have traditionally brought forth many more innovations than the giant US corporations that gobble up most of the federal industry research and development budget and spend most of the private money spent on industrial research in the United States

The "set-aside" is meant to do two things first to establish a federal policy that enhances the role of small businesses in research and development, second to generate more research results that can be turned into commercial activity on the open market The plan is modelled on a successful programme developed by the National Science Foundation that has enabled small firms to attract private capital, increase employment and, in a few cases, achieve great commercial success

Other countries are also trying to address this problem Great Britain has two smallish programmes, one that sets up "teaching companies", or long-term collaborations between universities and industry, and the CASE programme that works at the student level, placing individual students in companies in such a fashion that their industrial research earns them credit towards their degree Although having some direct federal programes in priority areas, West Germany relies primarily on tax incentives to stimulate industrial research and development Small German firms tend to contract with larger ones for the basic research they need to do before a project reaches the commercial stage All are concerned with the same thing how to transfer knowledge out of the basic research sector and into the increasingly competitive national and international marketplace

However, representatives of US universities have objected strongly to the bill, arguing that while the objectives are worthy, the mechanism of "set-asides" will shrink the share of the federal research budget that goes to the universities They would not care if Congress was in a spending mood and simply authorized more funds for small research and development firms But everyone in Washington is tight-fisted these days The Administration could not have supported the legislation, as it does now, if it proposed to add more money to the budget The legislation proposes to take the money from the existing budget, and so spokesmen for the universities have been crying, "Why us?", although the Senate version, which the House will probably also pass, limits the "raid" on basic research funds to one per cent of the basic research funds of any agency The objections from university spokesman are loud and clear However, there have been fewer objections from the contractors and applied researchers who now get the remainder of the federal research and development budget, some of whose funds will be used to make up the balance of the new programme

University scientists have reason to be concerned about protecting their share of the federal research and development pie from encroachments But the objections to this tiny set-aside — which is likely to be $300 million in all, of which a small part would come from basic research — have been strident even at a time when the workers at General Motors accepted voluntary wage and benefit cuts to help bail out that ailing enterprise If the most entrenched workmen's unions in the United States can make concessions to remedy the US economic troubles, cannot the better paid, supposedly more high-minded scientists?

In objecting to the legislation, spokesmen for universities and for basic research have been anxious to preserve their own turf at the expense of everyone else's Senator Warren Rudman, the prime mover behind the bill, is fond of quoting a prominent scientist at Massachusetts General Hospital stating he was not concerned about whether small businesses got more research and development, only about whether "one" of his investigators lost a grant because of it What is missing here is any perspective from the scientists that they are part of any larger effort Remember the old idea that the United States should invest in basic research for the sake of economic growth, productivity and the creation of jobs?

The new bills, therefore, put the US university science community squarely on the spot Thirty-seven years ago, Vannevar Bush and his colleagues persuaded the US government to sponsor basic research in the universities, and required that it have special, protected status, its own programmes and agencies, because, as Bush wrote, "Basic research leads to new knowledge It creates scientific capital It creates the fund from which the practical applications of knowledge must be drawn"

Small firms are also part of the "fund" from which practical applications are drawn, both those which do basic research and those which do not If the university spokesmen are truly concerned about national productivity, creating new inventions, and economic health — as they have said they were in selling their own budget requests to Congress in the past — they should support any measures that further that goal At least they should offer constructive alternatives But they do the image of science — and the US economy — no good by limiting their comment to paranoic fears that their particular sector will be hurt

0028-0836/82/180002 01$01 00

National labs reach crossroads

US report looks abroad for ideas

Washington

The Energy Research Advisory Board (ERAB) of the US government has issued its preliminary advice on a critical study of the fate of the Department of Energy's (DoE) national laboratories This is a key study because the laboratories are also being looked at by the White House, which may not be favourably inclined towards them and will have much to say about their future

At stake is the role of nine facilities (see below), some of which do the most intensive high technology research in the world — on atomic weapons, fusion and lasers — gobbling up \$2,300 million, or half of DoE's annual research and development budget each year Apart from their atomic weapons work, they are entirely the creatures of DoE Starkly put, the choice is whether the laboratories should throw their doors open to become national centres, performing research for other federal departments, private industry, perhaps even other countries, or whether, as budgets decline, they should concentrate on a few things they do well and consolidate, with some of them closing down

The debate will be joined in the autumn when the ERAB panel makes a final report At that time, the White House Science Advisory Committee will be studying the fate of the DoE laboratories and those of the Department of Defense Mr George A Keyworth II, the President's science adviser and an alumnus of Los Alamos, has said that he thinks some of the laboratories should close down Last December, the Administration proposed that all DoE research be moved to the Department of Commerce, a suggestion that no one in Washington now takes seriously And James Edwards, Secretary of DoE, surprised ERAB recently when he told them that, in his view, the laboratories should remain in existence and throw their doors open to become national technical centres So the discussion, once joined, will be lively

The ERAB panel is headed by Ivan Bennett, dean and provost of New York University Medical Center DoE requested the study last September A preliminary report was sought this spring to feed into the White House's review Although the group has not decided its own view of the laboratories' fate, the preliminary report outlines the laboratories' strengths and future options

Some of the institutions, like Los Alamos, were founded as part of the weapons programme and retain important roles in weapons research Others, such as Ames Research Laboratory run by the University of Iowa for DoE, do basic research Oak Ridge has undergone a major transformation into a general applied research laboratory since its beginnings during the Manhattan project

The appendix to the ERAB panel's report deals with how major laboratories in other countries have adapted to changing times or to the stark fact of having fulfilled their mission The UK Atomic Energy Authority's laboratories at Harwell are discussed at some length Founded in 1946 to assist in developing civilian nuclear power, Harwell's task was nearly complete by the mid-1960s Then, the report notes, Harwell kept the same disciplines but took on new clients outside the government's agency The lesson of Harwell, it states, is that "once a large laboratory has been assembled, it can be useful to government in fields that extend far beyond its original mission and that such organizations can indeed change their mission without loss of vitality "

National laboratories covered in the ERAB report

Ames Laboratory
Argonne National Laboratory
Brookhaven National Laboratory
Los Alamos National Laboratory
Lawrence Berkeley Laboratory
Lawrence Livermore National Laboratory
Oak Ridge National Laboratory
Pacific Northwest Laboratory
Sandia Laboratories

Other examples from abroad are Amersham International, the UK government's source of radioisotopes which has been spun off as a company, and France, where government laboratories commonly create companies to perform specific work for industry The report notes the Dutch pattern of numerous small laboratories, which are two-thirds government-sponsored and one-third sponsored by industry

Sweden's Studsvik Energiteknik AB, the government's energy laboratory, had three years' warning that its government funding would be cut by 70 per cent The three years enabled the group to make a smooth transition to becoming predominantly sponsored by outside interests

The implication for the US national laboratories is that they can adapt and so probably should

The carefully worded list of options presented by ERAB includes "concentrating" the laboratories' efforts through vigorous management by the DoE, freeing them of legal constraints so they can work for other sponsors, transferring other federal tasks to them, making them more multi-faceted or narrowing their mission "to correspond with resources allotted", closing one or more laboratories, or mergers The suggestions are intriguing, but to implement any one of them will be a gargantuan task

Deborah Shapley

Disarray on chemicals control

Brussels

Fears are growing in environmental circles that the special programme on the control of potentially hazardous chemicals set up by the Paris-based Organization for Economic Cooperation and Development (OECD), is running into difficulties as a result of the Reagan Administration's pro-industry and protectionist policies The downgrading of the US Environmental Protection Agency (EPA) has coincided with a loose implementation of the toxic substances control act (TOSCA) Jacqueline Warren of the Natural Resources Defence Council, a leading environmental pressure group, has cited EPA figures on the premarketing notifications for new chemicals submitted since April 1979 showing that 66 per cent of submitted notices failed to include toxicity data, thus undermining progress to establish procedures to reduce the risk of new chemicals

Officials at OECD's headquarters are cagey, in view of the United States' negotiating weight, about admitting allegations that the enthusiasm evident under the Carter Administration for the OECD programme has waned A legislative gulf has certainly grown between the United States and the EEC, as a result of the way that TOSCA is being implemented and the changes within the EEC after the Sixth Amendment came into force last September This is likely to impair prospects for rapid agreement within OECD

Under the Sixth Amendment of the 1967 directive on dangerous substances companies must submit a stringent pre-marketing dossier on a new substance not only to national authorities but also to the European Commission and to the other member states Not surprisingly, EEC countries are now disappointed that disagreements within the OECD still loom large in two crucial areas which appear to have been resolved by the 6th Amendment In October this year a high level OECD meeting will be tackling the areas in

0028-0836/82/180003-02\$01 00

question — how to ensure satisfactory quality controls for the implementation of the "Good Laboratory Practices" scheme agreed on last year, and how the confidentiality of data is to be guaranteed

The Good Laboratory Practices scheme represents a commonly applied definition of the standards and procedures to be used in laboratories testing new chemicals Unfortunately mistrust has clouded the agreement High standards cost money and governments are not convinced that everyone is abiding by the rules as strictly as they ought to be, so the October meeting will examine proposals for methods to check that the practices are properly upheld

This will be another small step towards the programme's ultimate aim of guaranteeing that the marketing authorizations handed out nationally are acceptable elsewhere Only by using the Good Laboratory Practices scheme can a standard dossier of data on a new chemical substance, the minimum pre-marketing data, be accepted by other national authorities

The scheme involves, though, a free flow of allegedly valuable commercial information, and this is causing worries in the chemical industry that such information will not remain confidential. The issue is complicated by the fact that there are large differences among OECD countries in the laws relating to freedom of information Environmentalists are equally worried that poor access to information on toxicity considered confidential by industry will violate the principle of public participation in the assessment of the hazards of chemicals to which the public are exposed

The OECD's agreement on exactly what information is relevant to hazard assessment has disappointed both the industrial and environmentalist lobbies The various OECD working groups have come up with a document which defines what data should not be considered confidential and which specifically excludes information such as spectroscopic data which might reveal the chemical identity of a substance before a manufacturer wishes it to be generally known A final agreement has yet to be reached in sensitive areas such as pharmaceuticals The United States is held to be particularly responsible for back-tracking on this question as the list of non-confidential data continues to be reduced If too much remains confidential, hazard assessment made by industry or government will be harder to question, especially if there is consistency on marketing authorizations

So it seems there is still considerable work to be done within OECD before guidelines are agreed on what hazards can be considered acceptable, and this may prove to be the biggest stumbling block to further international cooperation in this field **Jasper Becker**

How Dr Melamed lost his degree

In 1976, new regulations came into force in the Soviet Union governing VAK, the "Higher Attestation Committee" responsible for conferring the degrees of Candidate (PhD) and Doctor (DSc/DLitt) The full text was not published at that time (a booklet was promised for "later") but the media placed considerable stress on the fact that the new regulations demanded scrutiny of the political attitude of the postulant for the degree

It now appears that this requirement can work retroactively During the past year, at least seven Jewish scholars have been deprived of their degrees, for having exhibited "anti-patriotic" attitudes in applying to emigrate to Israel A document — which claims to be a transcript of such a deprivation process — has just reached *Nature*, and is reproduced here in a slightly abridged form **Vera Rich**

From the stenographic transcript of the Scientific Council of the Faculty of Geology of Moscow State University, 20 January 1982.

Chairman (Professor Adonis G Gayanov) Members of the board already know the agenda for today's session *inter alia* the stripping of Vladimir Grigor"evich Melamed of his academic titles due to his anti-patriotic dealings, unworthy of a Soviet scientist

Melamed is not present, he is ill On 18 January, two days before this session, he notified me that he had a sickness certificate (*reads Dr Melamed's notification*)

I feel it is particularly necessary to point out the problem of his absence for the following reason According to paragraph 5 of the regulations of VAK on the conduct of sessions for the stripping of academic titles, the person from whom the title is to be stripped has to be present

(*After some discussion, — including pressure from the public gallery — it was decided to proceed in Melamed's absence*)

A G Lyubimov Allow me to read the material from the special committee of the board, set up to investigate the activities of V G Melamed, who expressed the desire to emigrate to the state of Israel

> Melamed has been working at Moscow State University for approximately thirty years Latterly he was in charge of one of the laboratories, and was dealing with pressure-modification of the permafrost process He distinguished himself by his energetic attitude and was therefore admitted to the Party In 1975, he received the degree of DSc, specializing in geophysics
>
> In November 1980, he submitted a request to emigrate to the state of Israel, which is known to be capitalistic and no friend of the Soviet Union In view of the increasing severity of the international situation and all that has been stated previously, we find that the decision of Melamed is an anti-patriotic act, unworthy of a Soviet scientist
>
> We are of the opinion that Article 104 of the VAK on the stripping of scientific titles can be applied to him Dated 18 December 1981 Signed by all members of the committee

Chairman Who wants to comment?

Professor Epinat"eva I have a question about Melamed's scientific profile Would you please specify what he has achieved in the scientific field?

Chairman This is not the question We are concerned with something else, not science We are not going to discuss the scientific work of Melamed, so this is not relevant

Professor Dmitriev Will the minutes of the decision of the committee be added to the documents that VAK receives from us?

Chairman Yes

Dmitriev Then I want to point out an error Melamed did not defend his doctoral thesis in 1976, as the committee says, but in 1977 And it is an ill-chosen expression to say, "and all that has been previously stated" From this one might conclude that he should be deprived of his degree, not only because of his wish to emigrate from the Soviet Union but also because he worked for Moscow State University for thirty years

Chairman Good, we shall correct this Any further comments? (*Silence*) Good I call on Arkadii Vasilevich Kalinin, member of the committee

Kalinin Today our board has a rather difficult task, which is, however, completely justified by VAK regulations As you all know, Article 104 of the VAK regulations deals with the stripping of academic degrees due to conduct unworthy of a Soviet scientist This includes offences such as anti-patriotism This leaves our board one choice, to decide whether or not Melamed's offence can be interpreted as anti-patriotic or not

According to the law, emigration is permitted for the sake of reunification of families Melamed informed the departments which grant exit visas that this was his intention However, no law prohibits public organizations from judging Melamed's activities

Academic titles in the Soviet Union are conferred not only on the grounds of scientific criteria We consider the function of an academic title to be more important than in the capitalist world It confers not only the right to a function, but a salary If Melamed remained a doctor, he would continue to do things to which he no longer has a right His titles should be taken away from him

Professor Nikitin I should like to add something in order to show two aspects of the anti-patriotism of Melamed First, he was a Party member for a number of years and also a propagandist When submitting his papers applying for an exit visa, he turned his back on all his early ideas

0028-0836/82/180004-02$01 00

Second, he is a qualified specialist, a scientist Now he is leaving in order to work on behalf of Israeli science, to work for our opponents, our enemies And this is even more than anti-patriotism

Chairman Thank you Does anyone else wish to speak? The case is probably completely clear and we can begin voting

J V Medvedkov Permit me to say something, not everything has been said I am especially qualified to speak Before I became involved in soil-sciences, I studied law

It is, as you know, not only a moral judgement or the opinion of colleagues if Melamed is stripped of his academic titles they give a right to a scientific function, a salary, the possibility of scientific work appropriate to his academic level, admission to libraries etc

A case like this is treated completely unambiguously by a special juridical act I mean the Final Act of the Helsinki Accords of 1975 I should like to quote from *Izvestiya,* 2 August 1975

> The submission of an application for reunification of family shall not lead to changes in the rights and duties of the person who made the request or of members of his family

Hence it is inadmissible that the board should seek to limit Melamed's rights because of his wish for reunion with members of his family in Israel

I have not heard any proof that Melamed's conduct constitutes an anti-patriotic offence

It is no accident that in such a complete work of reference as the *Soviet Encyclopaedic Dictionary* (Moscow, 1979) the term "anti-patriotic" does not occur One can use an expression which comes close to it, namely "anti-Sovietism" This term does occur in the *Soviet Encyclopaedic Dictionary* On pages 63–64 it says that anti-Sovietism is an attempt to defame and misrepresent everything that the Soviet Union has achieved in economic, political and cultural fields

When it is stated that Melamed's conduct is anti-patriotic, then he is apparently also accused of anti-Sovietism

By whom and where has it been proved that Melamed has done anything to misrepresent and run down those things which the Soviet Union has achieved?

His application to emigrate was submitted without any publicity It is an act which falls within existing laws and standards He did not give public announcements and did not publish books which misrepresent and run down all that the Soviet Union has achieved

I fear that in fact, the Soviet Union will be harmed by those people who have applied to the board with the demand that Melamed be punished

Chairman Any further comments?

Mrs F Melamed My husband is ill at the moment and I do not understand how it is possible for his case to be considered *in absentia* Today a matter is being discussed that is of vital importance to him His entire 30 years career in the service of Soviet science has been put in doubt Can one call such a career unpatriotic? Why do members of the board disregard the signature of the Soviet Union below the Helsinki Final Act?

Professor Epinat"eva It is one thing if a worker at a normal research institute or even the Academy of Sciences expresses the wish to emigrate, but it is totally different for people working at the university to do so The latter are dealing with young people, aren't they? The act that Melamed performed distorts the world view of our youth

S A Kats This concerns the shattering of an entire human life And you limit yourself to two short sentences If the matter had been dealt with by the court, the procedure would have been very different

Professor Kalinin I am against the fact that "anti-patriotism" and "anti-Sovietism" are equated When Herzen left Russia, it was an act that we call "anti-Tsarism" and not "anti-patriotism"

It was the last straw to say that Melamed would be anti-Soviet But can we call his action complimentary to the fatherland? He did not want to leave Moscow State University in order to make life easier for us So we do not have to make life easier for him I do not think that we are in violation of the Helsinki Accords

G O Marshuhuk I myself disapprove of Melamed's action, and I told him so But disapproving is one thing and accusing him of anti-patriotism, and punishing him severely by denying him the right to scientific work is something totally different

I S Irlin There is no doubt that Melamed behaved within the framework of the law

During the past six to seven years, some thirty thousand people left the Soviet Union to be reunited with their families Among them there were quite a number with doctoral or candidate's degrees But personally I do not know of any case where one was deprived of his scientific degree In my opinion the board wants to create a precedent

Professor V V Karus According to Article 104 of the VAK regulations, one needs three qualities to hold a scientific title qualification, patriotism and regard for ethical standards Patriotism is love for the fatherland which raised us But Melamed decided to leave the fatherland at a time of severe international tension and for a country that propagates a malicious hostility against the Soviet Union The right to work will not be taken from him, but why should he get an extra salary for working in a scientific field?

Chairman I request you to vote for the stripping of Melamed's degree of Candidate of Physico-Mathematical Sciences For the motion? Carried unanimously I hereby declare the meeting closed □

Agricultural biotechnology

Soon to start

Springtime has not induced germination of the British Technology Group's (BTG) plans for an agricultural genetics company but it is still expected to bloom this year One snag at present seems to be the degree of scepticism among those of the calibre to be potential scientific directors about the ability of any such company to be profitable within five years, as BTG would want

Although BTG considers the topic too sensitive to discuss, by all other accounts the plan is that the new company should have the same kind of special relationship with the UK Agricultural Research Council (ARC) as does Celltech with the Medical Research Council (MRC) Celltech was also set up by BTG (or, more precisely by the National Enterprise Board which has since fused with the National Research Development Corporation to form BTG) and operates in the field of biotechnology Its relationship with MRC allows Celltech first refusal on any discoveries within the realms of recombinant DNA and cell hybridization that emerge from the various units of MRC In return Celltech channels money back into MRC and has an obligation to help the council find other commercial outlets for discoveries in which it has no interest

ARC and many of its employers who work on subjects that will be of interest to the new company (some of whom are already consultants to BTG) tend to be enthusiastic about the prospect for several reasons

The first is the possibility of an easy and financially benefical way for ARC to channel its discoveries of commercial potential in the domain of plant breeding At present it has no such outlet New strains of plants developed by conventional techniques within ARC institutes have to be given to the National Seed Development Organisation which, although a profitable concern, donates its profits to the Treasury rather than directly to ARC

For discoveries in gene cloning and transfer, that have some potential for being used in the genetic improvement of plant

0028-0836/82/180005-02$01 00

strains, ARC has no ready outlet but individuals have sometimes had to contend with the rather persistent advances of some of the several American companies that have been set up in advance of any British counterpart A special relationship with a company, preferably British, would therefore suit ARC well

For its part, BTG is believed to have board approval for starting the company and to have a managing director in mind It has not however, yet found a scientific director and has met with some refusals already

One problem facing any potential scientific director is the need to ensure that the company is profitable within five years That was also the BTG stipulation for Celltech but the problem is greater in plant breeding because the genetic manipulation of plants by modern techniques is far less advanced than is that of bacteria, which Celltech use to make products of commercial value, such as rennin

Nevertheless the optimists believe that profits could be made within five years by concentrating initial efforts on the improvement of bacterial strains, particularly the nitrogen-fixing *Rhizobium,* which are used to inoculate crop plants and on the development of techniques of clonal and meristem propagation

Finance for the new company is being arranged by BTG which is almost certain to provide at least one third of the money from its own, governmental coffers The rest will be raised from private sources

Peter Newmark

University admissions

Still squeezed

Applications from home and European Community students for undergraduate courses at British universities are likely to be six per cent up this year on last, according to figures released by the Universities Central Council on Admissions (UCCA) (see table) But admissions to courses starting in October 1982 are expected to be down on admissions in October 1981

This is no surprise The government has implemented its policy of limiting university places just when the number of 18 year olds in Britain is increasing Hence, the previous policy of gearing university places to demand has been abandoned The Department of Education and Science, in recent evidence to the House of Commons Public Accounts Committee, has thrown light on how this policy may affect potential students The table below shows the department's estimate of the numbers of potential students that may be deprived of a university place in the academic years 1981–82 to 1983–84

Understandably, universities hope to make up the fees lost from home students by taking in more students from overseas But their aim has been made more difficult since the government removed the subsidy from overseas student fees The figures clearly show that full economic fees have deterred a high percentage of potential overseas applicants

The number of overseas admissions, however, is not necessarily a constant proportion of applications UCCA statistics suggest that the shortfall in acceptances of places from overseas students in October 1981 over October 1980 was only 19 per cent, compared with the 34 5 per cent shortfall in applications But statistics compiled by the University Grants

Annual percentage change in applications for undergraduate courses at UK universities

	1980 over 1979	*1981 over 1980*	*1982 over 1981*
Home students	+3 5	+4	+6*
Overseas students	−12	−34 5	−20*

*Estimates

Committee (UGC) suggested an even smaller shortfall in overseas admissions only two per cent in 1981 over 1980 This apparent discrepancy seems to be explained by the fact that UGC includes admissions for more non-degree undergraduate courses than does UCCA So it seems that overseas students are now opting for shorter, less costly courses

Last week, the House of Commons Public Accounts Committee published the report of its findings on the administration of university grants The committee seemed pleased with the move by the education department to reduce the amount of university income not subject to cash limits by transferring the grant paid to home students for their fees directly into the universities' purse The universities are praised for keeping their student intake on target last October (a 4 per cent shortfall over the previous year) — but the polytechnics, and other institutions of higher education, come in for a drubbing for increasing their intake by 18 2 per cent over the previous year The committee's report urges that the body now being set up to control higher education outside the university sector be developed quickly with full co-operation of UGC and laments the fact that steps to coordinate all aspects of higher education had not been taken before the cuts to the universities

The parliamentary committee also seemed satisfied by assurances that the universities are taking care not to offer new tenured appointments with no redundancy clause The Committee of Vice-Chancellors and Principals has been looking at more flexible forms of contract

Judy Redfearn

No. of home students, aged under 21, entering university (thousands)

	*1980–81**	*1981–82*	*1982–83*	*1983–84*
UGC targets	67 3	65 2	63 1	60 9
Target to maintain 1980–81 age participation rate†	67 3	68 5	70 4	69 5

* Actual intake

† Age participation rate is the percentage of 18–21 year olds in the population entering university, which was 7 5% in 1980–81

US research support

Question of size

Washington

In coming weeks, the US Congress will probably pass legislation that would give a major shot in the arm to small research and development firms in the United States, many of which suffer in the present economic climate, yet which are major sources of new technical inventions The legislation is not final, however, and is subject to considerable opposition voiced by spokesmen for the universities and government-sponsored basic research

The National Science Foundation (NSF) estimates that there are 13,000 small firms in the nation, defined as independently owned firms with 500 or fewer employees and performing research and development work Although numerous, accounting for 85 per cent of firms involved in such work, these small companies in fact spend only 4 per cent of all industry research and development dollars In contrast, giants such as McDonnell Douglas and IBM spend 87 per cent of US industry's investment research

Yet there is ample evidence that most innovations come from small firms One 1976 study showed that small firms produce 24 times as many innovations per research and development dollar as large ones, even though the small firms receive only 2 per cent of total federal support for industrial research

Small business found a champion last year in Warren Rudman a freshman republican senator from New Hampshire Rudman introduced a bill that would set aside one per cent of all federal research and development funds — which total some $40,000 million — for small firms They would compete for the money by applying to separate federal agencies for grants, awarded on a peer review basis, as seed money If the work was fruitful, some firms would qualify for follow-on funding In a third phase, the money would have to come from the private sector, or from other government sources if the government was interested in the company's work

The plan was modelled on the Small Business Innovation Research Program run by NSF, and a newer, similar programme run by the Department of Defense The NSF programme gave some $5 million in seed money to 42 small firms in 1977 By 1981, the 11 of them that qualified for follow-on funding had

0028-0836/82/180006-02$01 00

succeeded in raising $41 4 million in outside capital and equity Moreover, the programme apparently created jobs The 11 firms had employed 261 people in 1977, by 1982 they employed 616 people The most spectacular growth was in a genetic engineering firm, Collaborative Research of Lexington, Massachusetts, which received $25,000 in seed money in 1977 and by 1982 had raised $24 9 million from outside sources

The Rudman bill made a spectacular passage through the Senate in December Of the Senate's 100 members, 85 were co-sponsors of the bill, and it passed by a vote of 90 to 0 One modification exempted the $10,000 million in-house federal research and developement from the calculation A second modification was an amendment introduced by Senator Harrison Schmitt limiting the amount of funds to be set aside that could be taken from federal basic research budgets This amendment was an attempt to placate spokesmen for the basic research community and univerisities who attacked the bill as a raid on basic research funds They argued that development work in most federal agencies has powerful protectors, whereas basic research does not In the Department of Defense, for example, the contractors and armed services buying the MX missile, or Trident submarine, would keep their research and development funds from the amount set aside, so that the basic research funded by the Department of Defense would be unduly tithed

One fear being raised by university spokesmen is that the small firms' share of the federal research and development pie will grow, at the universities' expense The proposed one per cent sounds modest enough, but any amount would take some funds away from federal basic research at a time when such money is becoming scarce

Some university spokesmen argue that small firms do not do basic research of high enough quality to qualify for federal funds, and that a set-side programme will allow them to adhere to this lower standard They argue that such firms should compete with universities and other traditional research groups Several federal agencies prohibit for-profit firms from applying for research grants, although the National Institutes of Health has now lowered this barrier

In the coming weeks the House will have to decide which version of the legislation it will pass The variant most palatable to university spokesmen is that proposed by Don Fuqua, chairman of the House Science and Technology Committee This would leave oversight of the programme to the authorizing committees of Congress for each of the federal agencies involved, thus allowing them to devise individual set-aside programmes or exempt the agencies under their jurisdiction

The version most likely to pass, however, is a bill put forward by John J LaFalce, which is modelled on the original Rudman bill but is even friendlier to small business The LaFalce version would make the money set aside not one, but three per cent of all federal research and development, and does not exempt federal in-house research from the calculation The LaFalce version would make $1,200 million available to the small firms in the first year — contrasting with the more cautious Rudman bill, which phases in the programme, reaching the $300 million level in the third year But in view of the opposition to the set-asides that has surfaced elsewhere in the House, it seems likely that the LaFalce forces would be satisfied with a final version limiting the set-aside to one per cent, having a three-year phase in period, and a feature limiting the "raid" on basic research

Deborah Shapley

Call from arms

Washington

At its annual meeting in Washington the National Academy of Sciences (NAS) made one of its rare ventures into public policy pronouncements The assembled members adopted a resolution calling on the President and Congress and their counterparts in the Soviet Union "to intensify efforts to achieve an equitable and verifiable agreement" limiting strategic arms, and to "reduce significantly the number of nuclear weapons and delivery systems" The resolution further urged them to reduce the risk of accidental war, to inhibit proliferation of nuclear weapons, and to "continue and observe" all arms control agreements including Salt II, signed by the United States and the Soviet Union in 1979 but not ratified by Senate Finally NAS urges the avoidance of "military doctrines that treat nuclear explosives as ordinary weapons of war"

The NAS resolution makes no mention of the "nuclear freeze" urged by other groups around the country It was passed almost unanimously, with a few abstentions and one dissenting vote Proposer for the resolution was Marvin Goldberger, president of California Institute of Technology and chairman of the academy's Committee on International Security and Arms Control The resolution was sent to the President via his science adviser, George A Keyworth II

Deborah Shapley

Polish Academy of Sciences

Slow progress

Poland's new legislation on the Academy of Sciences will ensure parity of funding for the institutes of the academy and the research institutes of the production ministries, according to Warsaw radio A main concern of Polish scientists has been the lack of separate budgets for the various institutes funded on the principle of dividing research into "problems" funded nationally The new bill, which is under discussion by the Council of Ministers (Cabinet), thus seeks to redress one of the major grievances of the academy scientists expressed at last September's National Congress of Solidarity in Gdansk It therefore forms part of a current tacit policy on the part of the ruling Military Council for National Salvation (WRON) to grant various "social" demands from the Solidarity programme while keeping open the question of the future of the independent trade union movement

Much, however, remains uncertain, and nothing has been announced so far about one of the most contentious issues — the status of the Secretary of the academy At present, the incumbent of this post holds quasi-ministerial rank, and is responsible in the first instance to the prime minister, not to his fellow academicians During "renewal", as part of the nationwide drive towards "self-governance", there were strenuous moves (headed by the academy lobby within Solidarity) to change this anomalous state of affairs and ensure greater autonomy for the academy, thus ending the long-standing friction between the members and scientific employees of the academy on the one hand, and the academy bureaucrats on the other

There has also been no news since the military council took power of many other proposed reforms, despite their apparent innocuousness For example, it was proposed that the academy should decide, on purely academic grounds, whether or not its members should be able to travel abroad At present, non-scientific criteria still play a major role in such decisions The emergency regulations on foreign travel for scientists stress that the would-be traveller must be given a thorough political vetting

Not surprisingly, this can pose problems for academy scientists A case in point is that of Artur Swiergiel, a young physiologist who, since last October, has been researching at the Babraham Institute of Animal Physiology in Cambridge

Mr Swiergiel had a six-month scholarship under an agreement between the British Council and the Polish Academy of Sciences Last November, realizing that he would need extra time for his experiments, Mr Swiergiel applied for an extension On 31 March, he received a telegram from Professor Maciej Zurkowski, director of the academy's Institute of Animal Breeding and Genetics, confirming the extension Three weeks later, a second telegram arrived, stating that Professor Zurkowski had been informed by "the academy" that the extension had been refused No explanation was given — but Mr Swiergiel had formerly served on the Warsaw regional executive of Solidarity

Vera Rich

0028-0836/82/180007 01$01 00

Muscular dystrophy research

Positive move

Not content with the paradoxical increase in public donations to charities in times of recession, the Muscular Dystrophy Group of Great Britain (MDG) is aiming to raise an additional income of £1 5 million from industry over the next four years And to coordinate the use of this money, earmarked for supporting clinical trials and work on the X-chromosome with that of their regular income, MDG have created the full-time post of research development director and filled it with an eminent scientist

At the age of 59, Professor Arthur J Buller takes on the new job In his career Professor Buller has moved across the spectrum of bodies supporting UK medical research, serving as a member of the Medical Research Council (MRC) from 1975 to 1981, becoming dean of the Faculty of Medicine at Bristol University in 1976 and moving to the government's Department of Health and Social Security as Chief Scientist in 1978 Now he feels less optimistic about public sector support for research and views the drain on university funds and the cutback in government support for research as a serious threat to the dual support system Medical charities can, he feels, ameliorate this situation

MDG works on an annual budget of £1 5–2 million raised through 450 voluntary bodies (the Muscular Dystrophy Association of the United States has a budget in the region of $16 million) In 1981 a total of £1,321,719 was made available for research — 45 specific projects have been approved for the academic year which began in October 1981 The biggest handout will go to the group's own research laboratory in Newcastle (£189,254) Two London hospitals will receive over £100,000 for work investigating tissue culture in dystrophic muscle and in the development and differentiation of single muscle cells

The additional money gathered by the fund-raising committee will be diverted to what the group describes as "significant new research developments" It is hoped that £1 million will go to work on the X-chromosome The techniques of recombinant DNA and gene cloning are already being applied to the X and other chromosomes with the dual aims of developing reliable methods of prenatal diagnosis and discovering the underlying genetic defect of several incurable diseases Such work is already funded by the MRC and the Cystic Fibrosis Trust but MDG want to make certain that efforts are specifically directed to Duchenne muscular dystrophy, a disease linked to an unknown defect of the X-chromosome

The other £0 5 million that MDG hope to raise from industry will be spent on clinical trials The assessment of these is likely to be helped by the availability in a London hospital of a £200,000 nuclear magnetic resonance machine This was purchased with the assistance of MDG and will enable muscle metabolism to be studied without taking tissue samples

MDG is a working example of how medical charities can give invaluable support to research when public sources are running dry It may also prove that charity can keep new research initiatives at home — additional funds help stem the brain drain

Jane Wynn

Musical chairs at the CEA

The French Commissariat a l'Energie Atomique (CEA) is to be reorganized, little more than a decade after the reforms of Andre Giraud set the CEA on its feet again

In Giraud's time, the problem was that the CEA had lost a major battle over the choice of power reactor for a French nuclear programme CEA wanted the indigenous French gas-cooled system, whereas Electricité de France backed the American pressurized water reactor (PWR) Electricite de France won, and the broken CEA had to be given new confidence Giraud achieved that, and also helped set up France's nuclear industry, now so successful, behind the PWR

Why the need to reform again? In part, CEA has outgrown the Giraud structure It now controls many subsidiary companies whose management needs to be rationalized But, more importantly, the new broom of the socialist government is to sweep clean The catchwords are decentralization and regionalism "Some reform of the CEA would have had to take place anyway" a CEA spokesman said last week "but it's going to go further with the new government than it would have done with Giscard d'Estaing"

M Michel Pecqueur, CEA's administrator-general, is to announce details of the reforms this week, with the hope — if the trade unions agree — that the budget for next year should be determined under the new structure That means things must move fast, for the budget process begins in June

One key issue will be the role of the seven *délégués*, who control the key sectors of CEA (nuclear applications, reprocessing, research and so on) Some say that these Paris-based posts carry too much power, and there must be some devolution to the units that the *délégués* control — among them the fundamental research laboratories at Fontenay-aux-Roses However, one of the délégues last week expected their roles to remain much the same after the reform The names of the posts might be changed, but the duties would not Indeed, it is claimed that the reforms will not be revolutionary No doubt that is CEA's intention

Robert Walgate

Ecology in the Sinai

Time of change

The final hand-over of the Sinai last week brought to an end an era of strict conservation and ecological surveying. Although there had been proposals, at the time of the Camp David talks, that the Sinai should become an "International Ecological Park for Peace", these have failed to materialize Moreover, according to Israeli ecologists, Egyptian plans to develop the Sinai could lead to the repetition of Israeli ecological mistakes committed with the best intentions by the early settlers

Under the Israelis, three major ecological centres had been established a bird migration monitoring station at El Arish on the Mediterranean, a field studies school near the Santa Caterina Monastery, and the Ras Muhammed underwater observatory None of these is still in operation, although a special agreement was signed ensuring the continuation of the Santa Caterina school The unique coral habitat at Rad Muhammed has been virtually destroyed by soldiers fishing with hand-grenades

Under Israeli control, ecological protection in the Sinai reached an almost ludicrous level, with tourists forbidden to pick up seaweed or coral from the beach for fear they might then drop it in the desert and upset the ecological balance

The Egyptians have ambitions plans for developing the Sinai, but Israeli ecologists know all too well the ecological hazards of the pioneering spirit — citing drainage of the Huleh swamps in North Galilee, and Ben Gurion's plans for intensive development of the Negev

Whether or not in cooperation with Israel, the Egyptians are matching Israeli developments They are particularly interested in developing solar ponds as energy sources, and Israeli plans to replenish the Dead Sea from the Mediterranean One Egyptian plan envisages an artificial "Dead Sea" in the Quattana depression, south-west of El Alamein As this lies below sea level, the canal bringing in Mediterranean water could be used to generate power, while the resulting salt lake could, as in Israel, be concentrated by evaporation to form the basis of a chemical industry

The emotional and political overtones of the return of the Sinai to Egypt make it unlikely that the Egyptians would seek Israeli advice on how best to develop the area However, the Israeli ecologists, who have worked long hours during the past few months to make their Sinai studies as complete as possible, express themselves willing, to make their data available to the Egyptians.

Vera Rich

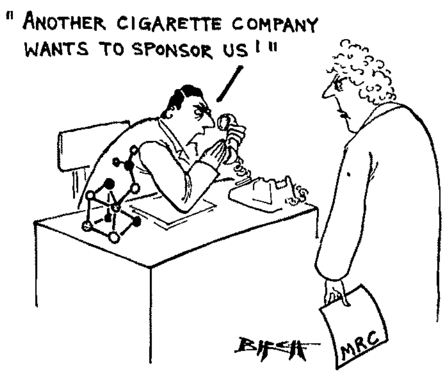
0028-0836/82/180008 01$01 00

MRC research unit

Job vacancy

The immediate future of the British Medical Research Council's (MRC) Pneumoconiosis Unit in South Wales seems assured, although life there remains unsettled Early last year, the unit had been the subject of a regular review by the MRC which had recommended staff cuts and greater emphasis on research supported by industrial contracts The fears of staff that the MRC was preparing to wind down the unit for final closure in 1987 when the director was due to retire, were heightened when the director, Dr Peter Elmes, left his post last January Those fears were allayed, however, when the MRC announced that it would appoint a new director — advertisements will be placed shortly In the meantime a four-person committee is in charge of the unit's daily administration

Dr Elmes's departure, however, highlights a problem for a handful of MRC units whose work is largely vocational Those units must win some of their money from contracts placed by the Health and Safety Executive, to whom the MRC still hands over some of its budget under the Rothschild customer/contractor principle Frustrations can arise when the contracts are for routine toxicological studies which divert staff from fundamental research for which the units are primarily intended



The MRC has used the problem to argue for an adequate budget to maintain fundamental research But it has also urged the units to win more contracts from industry as well as the Health and Safety Executive to help it out of its fiscal difficulties The more general problem, of course, is that routine toxicological studies on occupational hazards must be done and there are few suitable independent laboratories outside the MRC

The MRC's plans for the Pneumoconiosis Unit have begun to take shape over the past two or three years Hence there has been a shift in emphais from clinical to *in vitro* research and the unit's staff has been reduced Since the MRC's review early last year, the unit has received regular visits from a committee to advise on its work and research commissioned by industry has increased Dr Elmes, however, was not happy about the changes He mentions delays to research work caused by lengthy negotiations with industrial customers, both at the outset of a piece of research and during its progress, and industry's reluctance to publish results quickly

An illustration of some of the difficulties is provided by the case of kaolin workers in Cornwall The MRC, according to Dr Elmes, had been unwilling to support studies into the causes of an excess of lung disease in the workers on the grounds that they did not constitute fundamental research Industry and the Health and Safety Executive, which did come up with the money, awarded funds piecemeal for individual studies as the project progressed, thereby causing delay Publication of the final results, which apparently suggest a link between kaolin exposure and lung disease has been delayed because of industry's desire to reduce the risk before it is widely known

The MRC believes that many of these problems can be avoided if the terms under which research is conducted are made clear to industry at the outset Hence, the council would like the Pneumoconiosis Unit to continue its efforts to win more industrial contracts But the precise balance between applied and fundamental research, it says, remains to be seen One important factor in determining the unit's structure will be the views of the new director **Judy Redfearn**

US science education

Nearing a crisis

Washington

Mr George A Keyworth II, the President's science adviser, termed the "next few years" as "critical in science education in the United States" But he has extinguished any hopes that the Administration will do anything about it until the next presidential term

In a speech to the National Science Teachers Association (a group whose ability to carry out its job has been impaired by Administration cuts of all funds for pre-college science programmes run by the National Science Foundation (NSF)), Keyworth repeated the unfortunate truths many US leaders have learned of late about how poorly, and how little, science and mathematics are taught in US schools

For example, these subjects consume only seven per cent of elementary school teaching time, and even less time at the secondary level The ability of students in science and mathematics has deteriorated steadily, and most people leaving school and going into non-technical professions have little idea of what science and technology are about The public, he noted, spends $100,000 million annually to run US public elementary and secondary schools, while it spends around $6,000 million a year playing video games

Keyworth absolved the federal government of any role in solving the problem, except to produce graduate students in science "The alarm being raised about reductions in federal support for science education is misplaced We ought to be alarmed that those organizations and people spending the really big amounts of money seem to assign low priority to science education

"Our most important job is to work together to convince our local communities, school boards and universities that science is as basic as history, that students who must study English in the twelfth grade must also study maths But the federal government cannot and will not make this happen We parents, teachers and citizens must take up this challenge directly "

Rebutting this argument is F James Rutherford, who designed some of the high school science programme set up as a response to the Soviet Union's Sputnik, and who testified a week after Keyworth's speech to the opposite effect. He said that US elementary and secondary education is so decentralized and so locally controlled that no one is spending "really big amounts of money" There is no central focus for change or curriculum development except at the federal level, said Rutherford, who is now chief education officer of the American Association for the Advancement of Science

Keyworth had argued that the post-Sputnik science education effort simply did not take root and that geography and demography worked against it Rutherford says the decline was a direct result of earlier cuts in the NSF programme for science education (which once consumed half of NSF's entire budget) The Reagan Administration, for two years, has run down the remaining $80 million of that money so that only $15 million will be left in 1983 and this is allocated to graduate student fellowships

Rutherford notes that the timing of the Administration's initiatives precludes anything further happening during Reagan's current term as president The Secretary of Education, Terrel H Bell, has established a National Commission on Excellence in Education, due to report in October 1983, when the government's 1984 budget will be virtually complete John Slaughter, director of NSF, has said that his most important task as NSF director will be to set up another commission, on pre-college education in maths, science and technology, to report in late 1983 Thus these studies would have no impact, Rutherford notes, before the 1985 budget at the earliest, after the next election

So while Keyworth calls the problem critical, he does not propose to do anything about it, except to have the teachers lobby their communities, and weep

Deborah Shapley

0028-0836/82/180009 01$01 00

CORRESPONDENCE

Sequence libraries

SIR — Thank you for your article "Europe leads on sequences" (*Nature* 15 April, p 596) on our efforts to establish a nucleotide sequence library I welcome the attention given to the problems of data collection — a general discussion of these matters is both timely and important

I do feel, however, that a few points require clarification First, credit for the library at the European Molecular Biology Laboratory must also be given Kurt Stueber (of the Genetics Institute, Cologne), from whose collection we started, and who continues to work closely with us Similarly, any discussion of nucleic acid sequence data collections in Europe must mention Professor Richard Grantham's group at Centre d'Evolution Moleculaire (Lyon), who have run a large data bank for many years

Although a central data library has yet to be established in the United States (later this year according to NIH), several groups there have maintained large and excellent data bases for some time The largest of these are the collections of Dr Walter Goad and his group at Los Alamos Scientific Laboratory, and Dr Margaret Dayhoff and her group at the National Biomedical Research Foundation (Georgetown)

This is not a case of "too many data banks" The different approaches taken have each generated their share of good ideas, and if the present cooperative atmosphere can be maintained, a really excellent international resource combining the best ideas can result

GREG HAMM

European Molecular Biology Laboratory, Heidelberg, FRG

"Useless" research?

SIR — *Nature*'s leading articles show a persistent concern for the well-being of British universities, lately concentrating on research support and on the dual support system, at the expense of more important matters

We are all concerned to protect research but I believe you seriously underestimate the resilience of the British research effort and its supporters The latter are strong, vocal advocates of their cause and I know of no good reason why we should be alarmed at this moment Much of the fundamental research in question is expensive and in no way urgent It will survive a little slowing down and a little less international competition

Of much greater concern is the industrial-economic basis of the country and the contribution to that basis which needs to be made by universities and, of course, by other agencies of higher education In this context you mention (without, however, developing) the need for educational diversity It is an essential requirement of any successful organization that it posses the highest degree of diversity compatible with its integrity The model for a successful university system should be the usual model of any evolving system, namely a dynamic steady state stabilized in this case by the inevitable boundary conditions of budgets, standards and student numbers It is essential that, within the system, there be both lateral and longitudinal flexibility in subject content and student effort Such a pattern leads inevitably to the co-existence of broad degrees and specialized degrees, of vocational and non-vocational options, of universities and polytechnics, and of low level and high level exit qualifications There needs to be a marked change in the emphasis of university teaching towards the Design-and-Make Society and I would remind you of the Royal Society of Arts initiative in support of Education for Capability

There is nothing radically wrong with British universities that a little loosening up will not cure The present financial crisis, for example, will be coped with readily if all those over 60 or, if necessary, over 55 will simply make way for the young and for any subsequent changes in subject emphasis deemed necessary Making way does not imply "walking the plank", but rather continuing one's vocation in teaching, research or both in retirement or semi-retirement Some mortgaging of our financial future might still be necessary but not for long. Not every university can be saved an ugly confrontation in this way but the system as a whole could be Since I am in favour of the present number of universities, a little more even spreading of the costs and the cuts would be entirely in order Any suggestion that this would deprive us of necessary excellence is nonsense

In winning back a little of the country's confidence in us, it would be as well not to go on supposing that our present arrangements are the best of all possibilities A glance at France, Germany and the United States would convince many that this is not so

The single honours degree is not the be-all-end-all For its own good, the dominating influence of Oxbridge and its concentration on scholarly excellence and other platonic virtues will have to be challenged by other equally exacting paths of excellence, paths of university study which also lead to competence and usefulness If it is true that the annual toast of the Cambridge Philosophical Society is or was "Here's to our researches, may they always be useless!", is it surprising that our industries falter? This simplified analysis is no doubt open to criticism, but I shall be content if it stimulates discussion of ways forward rather than of defences of the *status quo*

GRAHAM HILLS

University of Strathclyde Vice-Chancellor, University of Strathclyde, Glasgow, UK

Research accountant

SIR — In an otherwise admirable article on current affairs and research at the Imperial Cancer Research Fund in *Nature* of 15 April 1982 (p 595), Robert Walgate is misleading in his interpretation of the accounts

The income for the year ended 30 September 1981 is correctly stated to be £17 million, but £12 8 million (75 per cent) was spent on running the laboratories and extramural units, and £803,389 (4 7 per cent) on appeals, whilst a further £2 7 million (15 8 per cent) was spent on supporting the Liverpool cyclotron and trials with interferon In addition we earmarked a sum of £2 5 million towards the cost of replacing the research laboratories that we shall lose when the Burtonhole Lane, Mill Hill premises revert to the MRC in 1986

A B L CLARKE

Imperial Cancer Research Fund, London WC2, UK

Not all cranks

SIR — In his review of Gardner's *Science Good, Bad and Bogus* (*Nature* 28 January, p.351) Sir Peter Medawar asks "was it not Cuvier who named a fossil ichthyosaur as *Homo diluvii testis* — man-witness-of-the-flood?"

No, it was not And to make matters worse it was actually the good baron who showed the beast for what it was a Miocene salamander Later in his review, Medawar states "Not all scientific nonsense is written by cranks though quite a lot of it is the work of scientists who asseverate upon difficult subjects of which they have no deep understanding" Perhaps together with Medawar's "electrician-eugenicists" and "astronomer-microbiologists" we should include the immunologist-palaeontologist? Perhaps this is too harsh for I have no intention to Shock (*ley*) — it is just that Medawar has given little quarter to the many, doubtless deserving, persons who have come under his fire in the past and I only wish to see him play his own game according to Hoyle

ANDREW FORESTER

Institute for Environmental Studies, University of Toronto, Canada

April Fool!

SIR — In your April Fool's Day issue, the article on pages 392–393 ends with the statement, "Long life is a fishy business indeed" There is more than long life that is "fishy" here You have been caught in a pseudonym web again, although this time indirectly The authors of the article from *Acta Gerontol (Prag)* are kidding somebody, as "Dlouhy-Zivot" means long life in Czech, and "Ryba" means fish Anyway, hyphenated names in Czech are rare

FRANK A PITELKA

University of California, Berkeley, USA

Creative energy

SIR — As a priest with a keen interest in astronomy and cosmology I have long pondered over the apparent contradiction to the second law of thermodynamics that is present in the standard big bang theory I therefore read the paper "A model for the cosmic creation of nuclear energy"[1] with much interest

Having attempted to exorcize the notion of a literal interpretation to the opening chapters of *Genesis*, I was amused to read in the above article, 'We find that the main creation of nuclear energy started around 10 s after the big bang, and most of the exergy was created during the first few minutes, 85 per cent during the first hour, and that the process was essentially completed during the first 24 h " Perhaps we also wish to add, "And there was evening and there was morning, one day "!

"The first day of creation, who can act rationally on such a day!" — Alexander Solzhenitsyn

GARTH BARBER

The Chaplaincy, City of London Polytechnic, London E1, UK

1 Eriksson, K E Islam, S & Skagerstam, B S *Nature* 296, 540–541 (1982)

Institutional biosafety committees and public participation: assessing an experiment

Diana B. Dutton* & John L. Hochheimer†

The most innovative aspect of institutional biosafety committees, responsible in the United States for local oversight of recombinant DNA research, is mandatory participation from outside the institution. A survey of Californian committees and selected national data reveals wide variability in committee structure and procedures. Public participation, although constrained in various ways, has been generally constructive.

THE regulation of recombinant DNA research has evolved at a pace almost as dizzying as that of the research itself. Only seven years ago, the scientific community voluntarily halted certain experiments while potential hazards were investigated. In 1976, the National Institutes of Health (NIH) issued regulatory guidelines mandating specific methods of physical and biological containment, and prohibiting high-risk experiments. Since then, regulatory controls have been steadily relaxed in light of evidence that risks may be less than initially feared. But some experiments are still considered potentially hazardous, and disagreement persists about whether the research poses long-term or low level risks[1-3].

Another major trend has been regulatory decentralization. In 1978, revised guidelines shifted primary authority for enforcement from NIH to locally-appointed institutional biosafety committees in order to simplify administrative procedures and to encourage local responsibility, although NIH continued to monitor committee decisions[4].

Effects of NIH guidelines

Since 1978, the authority of biosafety committees has expanded to the point where virtually no federal oversight remains. Greater discretion has also been delegated to individual researchers, less than 15 per cent of permitted experiments now require prior approval from the biosafety committee. Retrospective review enables the committees to monitor safety standards without impeding most experiments. A recent report by the Congressional Office of Technology Assessment calls the guidelines "a comprehensive, flexible, and non-burdensome way of dealing with the physical risks associated with recombinant DNA research while permitting the work to go forward"[1].

The 1978 guidelines also instituted significant changes in public participation in decision-making. NIH's Recombinant DNA Advisory Committee was broadened to include more individuals from fields outside the biomedical sciences, each biosafety committee was required to include at least two members not affiliated with the institution (often called "public" members) to represent community interests. Further, biosafety committees were required to make the minutes of their meetings available to the public upon request and were "encouraged" to hold public meetings whenever possible.

These provisions were adopted in response to charges by public interest groups and others that decision-making had been dominated by scientists and technical experts[5,6]. They evoked sharply divergent reactions. Many scientists were openly sceptical about the public's involvement in complex technical issues[7,8]. Some warned — recalling Lysenkoism — that it could lead to political repression[9]. Others saw a greater danger in the widening rift between science and society, and looked to increased public involvement in science to heal this rift[10]. The 1978 guidelines clearly struck a compromise between the basic changes proposed by critics[11] and the pleas of scientists to "quietly dismantle the whole hateful (regulatory) artifice" (ref 7). Yet, these provisions did offer the possibility of a direct public voice in decisions at the local level and, in this sense, launched an experiment in public participation in science policy.

This study assesses the success of that experiment based on a survey of biosafety committees in California and national data. We focus especially on public participation because this is the most innovative aspect of the regulatory system, as well as the most controversial. The findings suggest that lay members have played a constructive role on biosafety committees — although constrained in various ways — and that involvement has been generally worthwhile.

The time is ripe for such an assessment. In November, the Recombinant DNA Advisory Committee considered making all the NIH guidelines voluntary rather than mandatory, but instead left them mandatory while again reducing their scope and placing still more responsibility for monitoring on biosafety committees. (This recommendation was recently adopted by NIH.) Despite this expanding role, biosafety committees have never been systematically evaluated. Previous plans announced by NIH for a two-year national study met strenuous opposition from biosafety committee chairpersons[12] and have not been pursued.

The regulation of recombinant DNA research symbolizes, to many, the way that society will deal with scientific and technological innovations involving potential risks. For such issues, as Court of Appeals Judge David Bazelon has noted, the only measure of confidence possible may be in the process by which decisions are made[13]. Thus, extraordinary effort has gone into developing a regulatory framework for recombinant DNA research that could provide sound, legitimate decisions responsive to local concerns, the Office of Technology report calls this framework "a possible model for societal decision-making on technological risks". The evidence presented in this paper points to certain strengths and weaknesses, but should not substitute for a more thorough investigation.

Survey methods

Data on Californian biosafety committees were obtained from surveys of chairpersons and nonaffiliated committee members. Questionnaires were sent to each chairperson of the twenty Californian committees registered in June 1980 with the Office of Recombinant DNA Activities of NIH. Nineteen committees responded, including twelve in academic institutions, five in non-profit research institutes and two in private corporations. Questionnaires were also sent to all of the 48 nonaffiliated committee members asking about committee performance and their own roles, forty-five questionnaires were returned, a response rate of 94 per cent. Due to these small sample sizes, few differences are statistically significant. Those that are significant are noted.

National data are based on a transcript of the plenary session of a meeting of about 200 biosafety committee chairpersons and other representatives, held in November 1980[12], and on a brief survey completed by

*Department of Family, Community and Preventive Medicine, School of Medicine, Stanford University, and † Institute for Communication Research, Stanford University, Stanford, California 94305, USA

0028-0836/82/180011-05$01.00

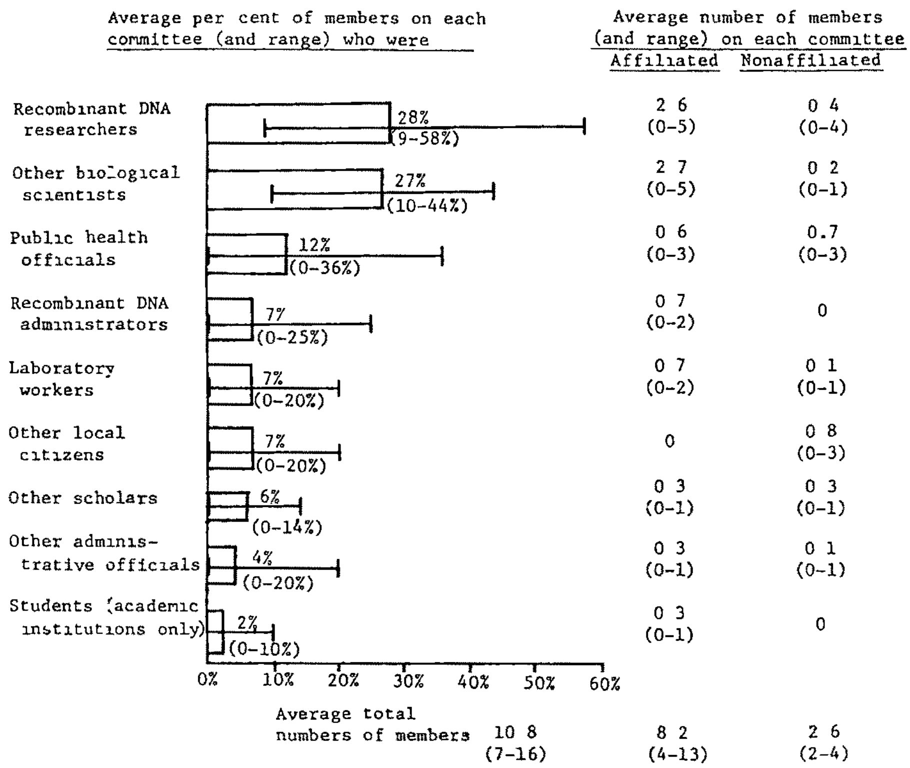


Fig 1 Membership of California biosafety committees.

98 of the 100 participants in the meeting's health surveillance workshop (survey results provided by S Barban, Office of Recombinant DNA Activities, NIH)

Survey findings

Since biosafety committees have had considerable discretion in interpreting guideline requirements, we describe first some general patterns of operation before discussing results on public participation

Most striking is the great diversity in almost every aspect of committee activities from the frequency and content of meetings to the performance of mandated functions For example, five of the Californian committees met once every three months during 1979, while several did not meet at all and one met fifteen times Most of these meetings were convened as the need arose, in fact, eight of the nineteen committees reported no regularly scheduled meetings About a third of the committees conducted some business by telephone or letter, several functioning this way almost entirely Some committees delegated substantial authority to chairpersons, others relied heavily on subcommittees Some committees were concerned only with recombinant DNA while others considered other potential biohazards

The number of research proposals reviewed also varied greatly Several California committees did not review any research proposals in 1979, while one reviewed 68 In both the Californian and national samples, about 60 per cent of all biosafety committees reviewed ten or fewer proposals Overall, committees spent about 30 minutes per proposal in meetings, but averages ranged from less than 10 minutes in one committee to more than an hour and a half in another In general, smaller committees reviewed fewer proposals and spent more time on each of them Some committees spent most of their time reviewing proposals while others concentrated largely on policy issues and other business In California, committees at private corporations devoted the least time to policy and committee business (20 per cent), while those in academic institutions devoted the most time to such matters (57 per cent)

The guidelines require that institutions ensure "appropriate training" for biosafety committee members, principal investigators and laboratory staff, conduct "appropriate" health surveillance of recombinant DNA research personnel; and adopt emergency plans for accidental spills[14]. But since specific criteria are not provided, committees have interpreted these responsibilities quite differently Nationally, the majority of committees offer no formal training courses Training for committee members in California varied from none to a combination of university courses, manuals and laboratory tours In the national sample, 89 per cent of the committees delegated responsibility for training laboratory staff to principal investigators, although several also assigned partial responsibility to the biological safety officer or to the biosafety committee

Plans for emergency spills also differed considerably Some Californian committees tailored plans specifically to recombinant DNA research, others relied on standard safety procedures In the national survey, 13 per cent of the respondents reported that their biosafety committees had no emergency plans for serious accidents or contamination, and 51 per cent reported no such plans for work-related illnesses Although serious accidents in recombinant DNA laboratories were thought to be rare, more than a third of the respondents said that their committees would probably not know if any had occurred About a third also said that laboratory staff were not told they were at increased risk of infection if pregnant or taking certain drugs Overall, 15 per cent of the respondents reported that their committees did not do a "good job" in protecting the health and safety of laboratory personnel

The majority of institutions had no special health surveillance for recombinant DNA laboratories In California, 13 of the 19 committees reported no such surveillance, for reasons ranging from "lack of evidence of medical hazards" to the statement that "routine monitoring procedures are adequate" Likewise, fewer than half of the committees in the national survey had "established medical programmes" Most industrial biosafety committees reported that employees were given standard comprehensive medical examinations modified only slightly for recombinant DNA laboratory personnel. Participants differed as to what form local surveillance should take (routine physical examinations and serological sampling were generally not seen as useful), but they did agree that the present uncoordinated approach would not allow detection of low level health effects.

Approval of research projects

The Californian biosafety committees approved 96 per cent of research proposals reviewed in 1979, 73 per cent of them without modification Only 4 per cent of all proposals were rejected, and the majority of committees (11 of 19) rejected none of the proposals considered

There was again variation among committees One, however, required modification of every proposal reviewed, while four approved all proposals without change Rejection rates varied within a narrower range (0–10 per cent) The two corporate committees approved the largest proportion of proposals as submitted, compared with those in academic and research institutions (86 per cent versus 76 and 61 per cent), required the fewest modifications (14 versus 19 and 36 per cent), and had the lowest rejection rates (0 versus 5 and 3 per cent)

High approval rates could mean that only high quality proposals were submitted, that proposals had already been modified through informal consultation, or that the committee's review was not

sufficiently critical. From the data to hand we cannot say which interpretation is most valid, although informal consultation appears to be common Most local decisions were eventually confirmed by NIH, since 1978 the Office of Recombinant DNA Activities has rejected only an estimated 5 per cent of proposals approved by biosafety committees (statement of W Gartland, Office of Recombinant DNA Activities, NIH, at 25 September 1980 meeting of the Recombinant DNA Advisory Committee)

Once proposals are approved, a number of committees apparently take no further formal responsibility for enforcing the guidelines At the national meeting, about a third of the committees represented in one workshop did not monitor laboratory procedures after initial protocol approval. Yet, as the workshop leader pointed out, problems are more likely after equipment has been operating for a while Because continuous monitoring is important, the guidelines include periodic review of experiments among the committees' mandated functions[15] Evidently this requirement has been widely ignored.

Composition of the committees

The guidelines leave the choice of committee members to the institution, but require each committee to include at least two non-affiliated "public" members, and to have a "biological safety officer" if research requiring P3 or P4 containment is being conducted[16] In California and elsewhere, most members are appointed by the administration In Cambridge and Amherst, Massachusetts, the city governments are also involved in selection, but such external involvement is unusual Most committees comply with the membership requirements, although the national survey revealed several institutions that had P3 facilities and no biological safety officer

The Californian committees ranged in size from 7 to 16 members, averaging 11, and all had the required number of nonaffiliated members Figure 1 shows the average composition of these committees, and reveals wide variation For example, recombinant DNA researchers comprised anywhere from 9 to 58 per cent of each committee, and there was comparable variability for most membership categories The "typical" committee consisted of 55 per cent recombinant DNA and other biological scientists, with five to ten per cent of the membership in each of the other seven categories Scientists were in the majority on most committees; only two contained fewer than 50 per cent scientists and one had more than 90 per cent Six of the 19 committees were composed entirely of men, and all were chaired by men Otherwise, there were few consistent patterns.

The guidelines state that the non-affiliated members "shall represent the interest of the surrounding community with respect to health and protection of the environment", and they list as suitable examples "officials of State or local public health or environmental protection agencies, members of other local government bodies, or persons active in medical, occupational health, or environmental concerns in the community"[17] Many of the Californian committees followed these suggestions. the two largest categories of nonaffiliated members were public health or other government officials (33 per cent) and local citizens (31 per cent)

The third major category reflects a rather different response 25 per cent of all nonaffiliated members were recombinant DNA or other biological scientists working at different institutions In two committees, the non-affiliated members consisted entirely of outside scientists. Although most such scientists were well qualified technically, their qualifications for representing community interests on public health or environmental issues were generally not evident from their *curricula vitae*

A major premise underlying the proposals to broaden participation in biosafety committees was that local citizens and other non-scientists would raise different issues from scientists, offering contrasting perspectives Data from the Californian survey support this premise. Predictably, lay nonaffiliated members had more trouble understanding discussions than did scientists (66 versus 43 per cent), and tended to rate their own contributions as less valuable Several lay members reported that their suggestions were given little weight. One commented that "the committee served mainly as a rubber stamp The tasks require specialized expertise "

On the other hand, lack of technical understanding could sometimes be advantageous. As one lay member said, it "requires the committee to think at a less hurried pace about potential problems . to think through problems from a somewhat different perspective", and to seek "information that a trained person might have assumed was implicit" One said simply, "I am the gadfly "

Scientists and non-scientists also had different views about community influence Lay nonaffiliated members were a good deal more sceptical than nonaffiliated scientists about direct community input to committee proceedings. Yet the nonaffiliated members, apparently because they viewed themselves as representatives of the community, were almost twice as likely as the scientists to say that committee decisions "always" took account of community views (38 versus 20 per cent)

Value of public participation

Committee representatives at the national meeting disagreed about the value of lay members Some thought that such members contributed little. Others claimed that community members played an important role — that they helped to avoid potential conflicts of interest and "caused the committee to operate in a more tidy fashion" Several stressed the useful political function of public members in providing a sense of open communication with the community and in defusing potential animosity

Do the benefits of public participation come at the cost of increased inefficiency or incompetent review? The Californian survery provides no evidence of such a tradeoff. Committees with local citizens reviewed almost twice as many research proposals in 1979 as those with no citizen members (20 4 versus 10 9 proposals), at the same time spending a somewhat larger fraction of meetings policy issues (36 versus 26 per cent) and covering more such issues (2 8 versus 2.5 per cent). Furthermore, although committees with citizen members spent less meeting time per proposal (27 versus 37 minutes), their review does not seem to have been more superficial than that of other committees, as they had a slightly higher rate of proposals rejected (6 versus 1 per cent) and of modifications requested (24 versus 21 per cent). These differences were independent of committee size, even though committees with citizen members tended to be larger, and larger committees tended to have lower approval rates Within both smaller and larger committees (Table 1), those with citizen members had lower approval rates as well as higher rates of modification and rejection.

With lay members in the clear minority on most committees, it is perhaps surprising to find even small differences. These findings are consistent with psychological research indicating that lay people tend to define technical issues more broadly than experts and to be more

Table 1 Comparison of rates of approval of research proposals for institutional biosafety committees with and without citizen members, by committee size

	Committee size			
	7–10 members		11–16 members	
Research proposal Approval rates	No local citizens	1–3 local citizens	No local citizens	1–3 local citizens
Per cent approved as is	70	63	98	81
Per cent approved with minor modification	29	35	2	7
Per cent rejected	1	3	0	11
Number of biosafety committees	5	7	3	4

cautious in assessing possible risks[18] And they bear out James Watson's warning that "public members may take regulation seriously, unlike the molecular biologists"[7]

Public access

Community representation can also occur by direct participation in committee meetings Thus, the guidelines "encourage" open meetings whenever possible[19]. Many commentators had urged stronger measures — mandatory open meetings that were publicized and held at convenient times — fearing that a voluntary provision would have little impact

Such fears were justified The Californian survey as well as the national meeting indicate that most committees have done little to encourage public participation; many have actively discouraged it Almost half (42 per cent) of the Californian committees held no open meetings Of the eleven committees that did have "public" meetings, only five held them on a regular schedule, in all cases on a weekday during working hours More significantly, these meetings were apparently not announced When queried about different forms of publicity, not one committee reported any announcement of meetings on bulletin boards, in campus or local newspapers or in other public media At the national meeting, most chairpersons also reported that their committees did not publicize meetings widely It is hardly surprising that public attendance has been minimal About half of the Californian committees holding "public" meetings reported that no one attended while others had typical audiences of one or two

As in 1978, many still dispute the value of open meetings Almost half of the committee chairmen in California felt that open meetings were not desirable Objections raised included the use of meetings as a "political" forum, possible violations of confidentiality, inhibition of frank discussions by committee members and fear that a lay audience ignorant of technical issues would impede the committee's operation

In committees that had open meetings, however, almost all the chairmen thought they were desirable, for reasons ranging from preventing public misconceptions to informing committee members about community views The performance of these committees compared favourably with those with closed meetings with regard to number of proposals reviewed and policy issues discussed (Table 2) And, like committees with public members, those with open meetings spent slightly less meeting time per proposal, yet were somewhat more critical in their judgements, rejecting more proposals as well as requiring more modifications These differences were independent of both committee size and citizen representation They suggest that open meetings do not make committees measurably less efficient and may, in fact, lead to a more critical review of proposals

Effects of open meetings

If there are differences between open and closed meetings, they probably have less to do with specific issues raised by audience members than with subtle changes in atmosphere Most people attending committee meetings asked general questions about research procedures or policies such as earthquake standards or laboratory safety Yet comments from nonaffiliated members suggest that open meetings may nonetheless have had an effect on the committee As one member noted, "community members are the only 'outsiders' at the meeting", adding, "Without them, who are the watchdogs?"

Assuming that open meetings facilitate direct communication with the community, one would expect committees with open meetings to receive more views from the community and to be more responsive Table 2 shows just the opposite pattern Nonaffiliated members on committees with open meetings more often said that the community had no input and that committee decisions did not always adequately account for community views One explanation may be that, as one member put it, open meetings "sensitize academic members to community concerns", leading to greater awareness of the ways in which the committee was not responding to the concerns they heard Alternatively, members of committees with open meetings may simply have had higher expectations concerning community influence Despite scepticism about the amount of community participation, however, almost all nonaffiliated members rated it as somewhat or very helpful

Minutes of meetings provide another source of public information about committee proceedings The guidelines require that such minutes and related documents be made available upon request[20] Here again, this requirement has had a negligible impact, only one of the Californian committees reported receiving requests for minutes Militating against such requests was the fact that the meetings were unannounced, and the minutes were typically stored in an administrative office not readily accessible

There is a potential conflict between public accountability and the privacy necessary to protect trade secrets, as the guidelines acknowledge So far, such conflicts have arisen mainly in proprietary institutions At the national meeting, corporate committees reported that meetings were not open to the general public Outside members, many paid by the companies, had to guarantee confidentiality, further limiting public accountability In California, only the two corporate committees reported restricting the agendas of meetings because of proprietary concerns, although four other committees took special precautions to protect potentially patentable results before releasing minutes

Are biosafety committees worthwhile?

Most chairpersons at the national meeting seemed to think that the time and effort spent by biosafety committees were greatly out of proportion to the risks of recombinant DNA research and therefore that the committees were largely unnecessary Opinions expressed in the Californian survey were strikingly different More than 80 per cent of the chairmen thought the guideline requirements concerning biosafety committees were "about right", while 91 per cent of the nonaffiliated members agreed that "biosafety committees as they presently function are worthwhile" This difference in views may be due not only to changes in perceptions of risk between 1979 and 1980 but also to the different sources of data — individual questionnaires versus the group process that produces a consensus

Comments in the California survey, as in the national meeting, highlighted functions

Table 2 Selected comparisons of biosafety committees with open and closed meetings

	Committees that have No open meetings	Committees that have 50-100% open meetings
Average number of research proposals reviewed in 1979	14 1	18 0
Average meeting time spent per proposal (minutes)	56	51
Proposal approval rates		
approved as is	79%	70%
approved with modification	20%	24%
rejected	1%	6%
Average number of policy and procedural issues discussed that were listed by chair	1 6	3 5
Per cent of nonaffiliated members who said community had no input into committee proceedings	38%	58%
Per cent of nonaffiliated members who said committee decisions did not always account adequately for community views*	47	78
Number of biosafety committees	8	11
Total number of nonaffiliated members	18	27

*P « 0 05

of biosafety committees Many nonaffiliated members believed that the committees had an important role in promoting safety, for instance, one commented that they "force researchers to meet strict standards, as enforced by the campus environmental health officer Several hazardous projects have either been aborted or redesigned because of our action" Others thought that biosafety committees provided a useful forum for internal review of controversial issues and also "a sense of security to the public agency and to the scientific community — no secrets, open communication"

Discussion

In combining technical evaluation of scientific issues with provisions for accommodating social values and limiting conflicts of interest, institutional biosafety committees represent a significant policy innovation The experience of the Californian committees — which appear to be typical of committees elsewhere — suggests three general conclusions about their performance

First, the considerable diversity of biosafety committees appears to reflect not only varying local circumstances but also variable effectiveness Furthermore, the varying approaches (if any) to surveillance prevent development of a standardized system that might be more capable of establishing safety or detecting hazards

Compliance has been variable even on specific requirements of the guidelines Not all institutions conducting P3 research have appointed biosafety officers, and many do not monitor experiments after initial approval Continuous monitoring is also hindered by infrequent committee meetings The Office of Technology Assessment reports that biosafety committees "usually meet monthly", but only one of the 19 California committees met that often in 1979, and the frequency of meetings has undoubtedly declined since then Stanford's committee, which met four times during 1979, now meets only once a year unless special problems arise (interview with D Perkins, chairman of the Stanford Biosafety Committee by M London)

The low rates of rejection of research proposals also raise questions about effectiveness While these low rates may result from high quality proposals, they may also be due in some cases to inadequate review by the committee Previous studies of similar institutional committees monitoring human subjects research, suggested that their low rate of rejection of proposed experiments was at least partly due to poor performance

A second conclusion that may be drawn is that, given the highly decentralized nature of the biosafety committee system, clear minimum standards for essential elements of the system would be helpful One of the key experimental variables in these committees has been public participation, yet the guidelines define this variable only very generally through the requirement for nonaffiliated members The stated intent is that such members shall represent the surrounding community with respect to health and the environment, and the regulations stipulate that institutions are expected "to adhere to the purpose of the guidelines as well as to their specifics"[21] But precisely *how* community interests are to be represented is left notably ambiguous, inviting disparate interpretations Thus, a quarter of all nonaffiliated members were biological scientists from other institutions, many engaged in recombinant DNA work and with no evident qualifications for representing community interests

More specific provisions concerning public representation, such as those recently adopted by the Food and Drug Administration for human subjects committees[22], could have avoided this ambiguity Provisions had been proposed in 1978 that committees include a certain proportion of non-scientists or reflect the demographic composition of the community, to ensure direct community involvement (following the examples in the guidelines) rather than representation by outside scientists[11]. But NIH rejected these proposals on the grounds that the biosafety committee "is in large part an expert committee whose essential function is to evaluate research protocols in respect to containment levels, using the explicit instructions of the guidelines Rigid quotas are not necessary"[4] With such ambivalence about the role of public members, it is not surprising that the regulations remained ambiguous

It would also have been useful to define what constitutes "appropriate" training for biosafety committee members Insufficient technical instruction compounded the frequent difficulty that many nonscientists had in understanding discussions, and a number stressed the need for more systematic training on basic research techniques and terminology

The role of the public

Minimum standards for critical factors such as technical support and provisions for community representation need not restrict local experimentation in meeting or exceeding these baselines Such standards are vital in a decentralized regulatory system where effectiveness depends largely on local structures and procedures Without mechanisms for assuring accountability to community interests, public participation in biosafety committees has not been fully tested

Yet it is clear that some committees did make a genuine effort to involve local citizens and were at least partially successful The results of these experiences suggest a third conclusion which, although tentative, is perhaps the most important given a chance, public participation seemed to work fairly well Most community representatives took their role seriously and contributed to effective committee performance by broadening discussion and encouraging more critical scrutiny of research proposals Public meetings, even with limited attendance also appeared to give committees a better understanding of community concerns without impeding committee operation Many members believed that the committees were an important channel for public communication about an issue that remains sensitive, and that public involvement caused no evident harm to the research or to science

Even if federal regulations are eliminated, the controversy over how to regulate recombinant DNA research is not likely to disappear It has recently emerged anew in communities where genetic engineering companies have been set up. Because NIH regulations do not cover commercially-sponsored research, several towns and states have recently passed or drafted ordinances extending the present guidelines to industrial research and adding other requirements[18] Thus, biosafety committees or their equivalent may well continue to be a key element in local regulation As a workable mechanism for direct public participation, these committees have established a significant precedent in the decision process concerning science and technology

We thank Drs William Gartland, Stanley Barban and Elizabeth Milewski of the Office of Recombinant DNA Research Activities, NIH, for making records available, Nancy Pfund and Ralph Silber for helping design and conduct the California survey, and Natalie Fisher, Valerie Herman and Kathy McFadden for technical assistance Funding was provided under a grant from the NSF's EVIST Program and the National Endowment for the Humanities

1 *Impacts of Applied Genetics Micro-Organisms, Plants and Animals* (US Government Publications Office, Washington DC, 1981)
2 Thomasson, W A *The Bulletin of the Atomic Scientists* December, 26-32 (1979)
3 Wright, S *Environment* November, 34-42 (1979)
4 *Federal Register* **43**, No 247 60108-60128 (1978)
5 Callahan, D *Hastings Center Report* April, 20-23 (1977)
6 Wright, S *Environment* May, 6 (1978)
7 Watson, J D *The New Republic* January, 12-15 (1979)
8 Davis, B D *Science* **186**, 309 (1974)
9 Wade, N *Science* **197**, 348 (1977)
10 Toulmin, S *New York Times*, 12 March (1977)
11 *Recombinant DNA Research* Vol 4, August–December 1978, Appendices (US DHEW, Publ No 49-1876, Government Printing Office, Washington DC, 1979)
12 Krause, R M *Edited Transcript of Plenary Session III of the November 24-25, 1980 Institutional Biosafety Committee Chairpersons Meeting*, 29 December (National Institute of Allergy and Infectious Diseases, NIH, 1980)
13 Bazelon, D *Science* **205**, 277 280, July (1979)
14 Sections IV-D-1-g, IV-D-1-h, and IV-D 3-d, respectively
15 Section IV-D-3-c
16 Section IV-D-4
17 Section IV-D-2-a
18 Fischoff, B , Slovic, P & Lichtenstein, S in *Progress in Resource Management and Environment Planning* Vol 3 (eds O'Riordan, T & Turner, R K) (Wiley, New York, in the press)
19 Section IV-D-2-g
20 Section IV-D-2-h
21 Section IV-A
22 *Fed Register* **46**, No 17, 8977, Part IX, pgh 56, 107(c), January 27 (1981)

NEWS AND VIEWS

New frequency standards from ultra-narrow Raman resonances?

from Peter Knight

THE high-resolution study of atomic and molecular resonances over the past decade has led to the establishment of an atomic time standard based on the very stable atomic beam microwave transition between the hyperfine ground states of the caesium atom. Laser excitation of atomic or molecular transitions offers potential improvements in this 'standards' field but has been dogged by linewidth problems The lasers jitter in frequency and the optical transitions investigated have been broadened by motion, collisions or by spontaneous decay, so that the gain in frequency over the microwave transition is offset by the increase in linewidth Much effort is being expended to produce more stable lasers and to generate narrower spectral lines, in part because of a general spectroscopic interest and in part to find new laser-driven atomic clocks that could supplement or even supplant the Cs clock This could meet new demands for ultra-precise time intervals A recent experiment reports a new ultra-high-resolution laser spectroscopic excitation which not only eliminates laser jitter problems but also almost all line-broadening mechanisms, including that of a finite lifetime, and leads to a spectral line whose centre is stable to great precision

Ezekiel and his colleagues at MIT and the Rome Air Development Center have measured an optical atomic resonance of width 650 Hz (Thomas *et al. Phys Rev Lett.* **48**, 867; 1982). This width is four orders of magnitude less than the inherent natural width of one of the states involved and almost certainly less than the linewidths of the phase-jittering lasers used to excite the transition This remarkable example of a 'subnatural' linewidth has major implications for frequency and atomic clock standards Already, the preliminary data are claimed to compare well with the short-term stability of a conventional Cs atomic beam clock Ezekiel's

Peter Knight is in the Optics Section, The Blackett Laboratory, Imperial College, London SW7 2BZ

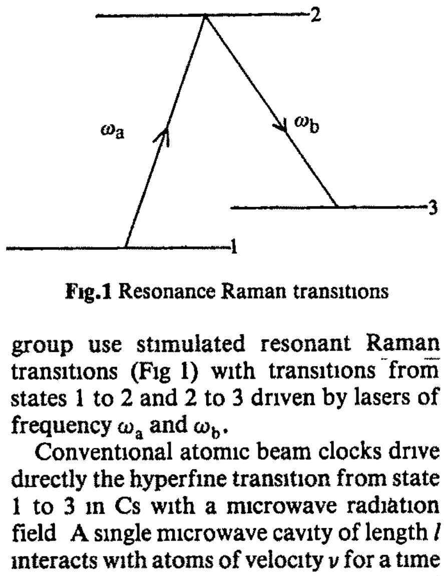


Fig.1 Resonance Raman transitions

group use stimulated resonant Raman transitions (Fig 1) with transitions from states 1 to 2 and 2 to 3 driven by lasers of frequency ω_a and ω_b.

Conventional atomic beam clocks drive directly the hyperfine transition from state 1 to 3 in Cs with a microwave radiation field A single microwave cavity of length l interacts with atoms of velocity v for a time $t \sim l/v$ as they fly through in a beam, and will excite a resonance at frequency $\omega \sim \omega_{31}$ ($\equiv(E_3-E_1)/\hbar$) of width of order $\Delta\omega = t^{-1} \sim v/l$ by simple Fourier arguments States 1 and 3 are both ground states and are not broadened by any other influence (there is no Doppler broadening in the microwave–beam interaction, there are no collisions in the dilute beam and no natural broadening of the line) To decrease this transit-time width $\Delta\omega$ and increase the resolution $\Delta\omega/\omega_{31}$ suggests increasing l In fact, two separate microwave cavities separated by a distance $L \gg l$ and fed in phase are used The atom has a probability amplitude a_1 of being excited in the first cavity and a_2 in the second The transition probability $P = |a_1 + a_2|^2$ has a two-slit interference pattern due to the quantum interference between different routes from indistinguishable initial and indistinguishable final states The interference lineshape is called a Ramsey resonance and has a width $\sim (v/L) \ll (v/l)$, making possible ultra-high-resolution spectroscopy

An optical transition between the ground state 1 and excited state 2 is not often limited by this transit-time broadening Instead the linewidth is produced by grossly larger influences The Doppler motion-induced spread in frequencies amounts to several GHz, which can be circumvented either using the new non-linear laser spectroscopic tricks, or by exciting a well collimated atomic beam perpendicularly with a well collimated tunable laser beam Providing the atomic beam is not too dense, collisional widths can also be eliminated The excited atom still interacts with its environment· it couples to the vacuum-quantized radiation field and consequently decays spontaneously with a lifetime typically of order 10^7 Hz This width seems immutable, and the spectral structure in the natural width forever unresolvable However this is not the case there are several ways in which atomic coherence can be exploited to generate subnatural linewidths, and I have reviewed these elsewhere (*Comments atom molec Phys.* **10**, 241; 1981) Nevertheless, we would not expect more than an order-of-magnitude improvement from even the most accomplished subnatural spectroscopist, simply because the penalty paid for the decrease in linewidth is exponential loss in signal strength

Ezekiel *et al* have used stimulated resonance Raman transitions from state 1 to 3 via state 2, by the absorption of a photon of frequency ω_a from one laser field and the stimulated emission of a photon of frequency ω_b into another laser field, such that $(\omega_a - \omega_b) \approx \omega_{31}$ A peculiarity of such transitions is that the Raman interference line has a width governed by the sum of the widths of states 1 and 3 only, with no real contribution from 2 Proof of this requires extensive analysis of the dynamics of the laser excitation and is true only for Raman difference frequencies $(\omega_a - \omega_b)$ close to ω_{31}; otherwise 2 will strongly contribute to the transition dynamics If the frequency ω_a is chosen to be exactly resonant with the ω_{21} transition frequency and ω_b is scanned through the ω_{23} resonance frequency and the decay fluorescence from state 2 measured, a Raman lineshape is generated in the fluorescence intensity versus ω_b Its overall shape is a Lorentzian of natural

0028-0836/82/180016-02$01 00

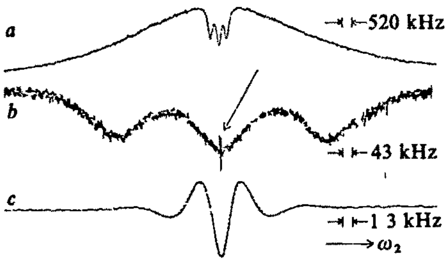


Fig. 2 Raman and Raman–Ramsey fringes observed by Ezekiel *et al* in sodium (from *Phys Rev Lett* **48**, 867, 1982) The arrow points to the fringes See the text for details

width (usually 10^7 Hz), but when $(\omega_a - \omega_b) \approx \omega_{31}$ (that is, at the 'two-photon' resonance frequency) a narrow non-absorbing 'hole' (G Orriols *Nuovo Cim* **B53**, 1, 1979) is burnt into the lineshape At this hole frequency, the atom is entirely transparent. Physically the Raman (or two-photon) hole is due to atomic coherence The atom is optically pumped by the two lasers into a coherent superposition of ground states 1 and 3 in which dipole moment induced on the 1–2 transition exactly cancels the 2–3 induced dipole This vanishing total dipole moment leads to zero net absorption and consequently no decay fluorescence The dipole moments only cancel close to Raman resonance elsewhere population can be excited into 2 and contributes a fluorescent, naturally broadened line

Ezekiel *et al* chose to study resonance Raman transitions in a sodium atomic beam in which state 1 is the $3^2S_{1/2}$ $(F=1)$ hyperfine ground state, 3 is the $3^2S_{1/2}$ $(F=2)$ hyperfine state and state 2 is the $3^2P_{1/2}$ $(F=2)$ hyperfine sublevel, of lifetime 16 ns The atoms are excited by co-propagating laser fields, and Raman dips are observed on a 10^7 Hz-wide fluorescent background when $\omega_a = \omega_{21}$ and ω_b is scanned through ω_{23} Three Raman dips are observed corresponding to Raman transitions between different magnetic sublevels in states 1 and 3, separated by the 300 mG Zeeman field applied to the atoms Each has a width governed principally by the transit time $t \backsimeq l/v$ in the region simultaneously irradiated by the two laser beams The two laser fields need to be correlated in their phase fluctuations or the two-photon, or Raman coherence will be washed out Ezekiel's group eliminate this phase fluctuation problem by deriving one frequency, ω_b, by acousto-optic modulation from a stabilized laser of frequency ω_a, such that phase-jitter in the two beams is precisely correlated Since Raman coherence depends in part on the phase difference between the light fields, the correlated jitter leads to a non-fluctuating Raman phase factor and to a Raman dip which can be, in principle, narrower than the laser bandwidth

As the Raman, or non-absorbing, dip is uninfluenced by the upper state decay lifetime τ, a Ramsey double-excitation technique can be used with a long delay $T \gg \tau$ without loss in signal but generating ultra-narrow lines similar to those in the Cs atomic clock The two correlated frequencies ω_a and ω_b are mixed and used to excite the sodium atomic beam in two regions separated by a distance L Figure 2 shows the sequence of lines observed for $L \backsimeq 15$ cm In Fig 2*a*, three Raman dips are observed on a 10^7 Hz-wide naturally broadened fluorescence line as ω_b is scanned though ω_{23} excited in a single interaction region A Raman width of 50 KHz is attainable Figure 2*b* shows the expanded scan of the Raman dips, now excited by two interaction regions separated by 15 cm, with central fringes corresponding to Ramsey interferences on the central Raman line (Ramsey fringes on the other dips are washed out by stray magnetic fields) Figure 2*c* shows the further expanded scan of these central Ramsey fringes Increasing the interaction separation L to 30 cm produces fringes of width 650 Hz (HWHM)

If these Ramsey–Raman techniques were applied to the Cs hyperfine Raman transition, a short-term stability $\Delta\omega/\omega \backsimeq 2\ 5 \times 10^{-11}$ seems possible over 1 s, which approaches the stability of a 30 cm interaction–separation Cs atomic clock The laser technique can be improved by decreasing laser-induced Stark shifts, optical pumping signal enhancement, laser-cooling of atoms to increase flight times and perhaps by applying it to ions in electromagnetic traps The new technique has arrived just as metrologists all over the world are seriously investigating new possibilities to replace the ageing Cs microwave atomic clock, and will be scrutinized carefully as a potential time-keeper of great stability □

Enhancing elements for activation of eukaryotic promoters

from Moshe Yaniv

TINKERING with recombinant DNA systems for the expression of genes in eukaryotic cells has recently thrown up a new class of regulatory sequences These are the 'enhancers', short viral sequences that can stimulate, by up to two orders of magnitude, the transcription of coding sequences from their own promoters Even more remarkably, they do so whatever, within reason, their position or orientation with respect to the coding sequences.

Enhancers were first discovered as sequences near the early genes of simian virus 40 (SV40) and polyoma that are required for virus viability[1-4] Approximately 30 base pairs (bp) upstream of these genes is found the TATA sequence typical of eukaryotic genes transcribed by RNA polymerase II and which appears to determine the specific site of initiation of transcription In SV40, between 113 and 257 bp upstream of the RNA start point is a directly repeated 72 bp sequence Removal of one 72 bp sequence has little effect, however, if part of the second sequence is removed the genome is rendered non-viable because at least one of the repeats is required for transcription of SV40 early genes

It is now clear that viral enhancers can also exert their effects on cloned nuclear genes Schaffner's group has shown that the SV40 72 bp repeat can increase the expression of a cloned rabbit β-globin gene in HeLa cells by 200-fold The globin gene was introduced into the HeLa cells by the calcium phosphate co-precipitation technique and its expression was measured after 2–3 days, before the integration of this DNA into the cellular genome The SV40 72 bp repeat has to be covalently joined to the gene (that is, it functions in *cis*), but surprisingly it works even if it is several kilobases upstream or downstream from the transcription initiation site, and its orientation with respect to the gene is unimportant[5] Similarly, Chambon's group has shown that the activity of the chicken conalbumin or adenovirus major late promoters can be stimulated by the SV40 enhancer[6]

These observations clarify and extend earlier suggestions that viral sequences may affect the activity of eukaryotic genes Capecchi's finding that the transformation of TK⁻ (a thymidine kinase mutant) mouse cells by plasmids carrying herpes virus *tk* gene is more efficient when the plasmid also carries the SV40 genome was probably mistakenly interpreted in terms of more efficient integration[7] The stimulation of expression of rabbit β-globin gene in HeLa cells by a 244 bp DNA fragment from polyoma virus[8], of human β-globin and in certain cases of the herpes *tk* gene by the SV40 72 bp element[4,9], and of the chloramphenicol acetylase gene linked to chicken α-collagen promoter by SV40 sequences [10] are probably other demonstrations of enhancers in action

Enhancers do not seem to be limited to papovaviruses 72 bp directly repeated sequences are also found in the long terminal repeats that flank integrated retroviruses (for example Moloney sarcoma virus) Although these sequences have no obvious sequence homology with

Moshe Yaniv is at the Institut Pasteur, 28 rue du Dr Roux, 75724 Paris Cedex 15

0028-0835/82/180017-02$01 00

those of SV40 they can replace them to produce viable hybrid molecules[11]

The presence of enhancers in retroviruses may explain some recent intriguing findings that appeared to be inconsistent with the promoter insertion model of viral carcinogenesis[12] According to this model, integrated retroviruses can provide active promoter sequences that increase the expression of cellular genes. By integrating upstream of the cellular homologues of viral oncogenes, and consequently increasing their expression, retroviruses lacking their own oncogene are thought to be able to transform cells However, Payne *et al.* have shown recently that in bursal lymphoma cells, avian leukosis proviruses can be integrated upstream or downstream and in either transcription orientation relative to the cellular *myc* gene (*c-myc*) — the cellular homologue of the putative transforming gene (*v-myc*) of myelocytomastosis virus Clearly these findings are inconsistent with a simple promoter insertion model, however they fit nicely with the known properties of enhancers

An interesting possibility is that enhancer sequences may have a role in regulating cell differentiation Polyoma viruses are normally unable to grow in early embryonal mouse cells, or in differentiated derivatives of these cells Mutants of polyoma with the capacity to replicate in embryonal cells were found to have sequence rearrangements[13-16] in the region of the virus shown to contain enhancing sequences, although not of the 72 bp repeat type[8] These results suggest that enhancers may be involved in tissue- or cell type-specific control of gene expression

The mechanism of action of enhancers is unknown but several possibilities may be envisaged Prominent amongst these are that enhancers may act as chromatin organizers, RNA polymerase entry points, sites of attachment to the nuclear matrix or regulators of local DNA superhelicity Of course these possibilities need not be independent or mutually exclusive Some evidence that enhancers may regulate chromatin structure comes from their presence in regions of papovavirus genomes not packaged into nucleosomes in cells[17-19] This may lead to changes in the structure of chromatin propagated on both sides of the enhancer and thus permit efficient transcription of neighbouring genes Alternatively the nucleosome-free region may act as a high-affinity site for binding RNA polymerase[20]

How enhancers really work, and whether they have a role in the normal control of cellular gene expression, are clearly exciting problems likely to be the subject of much future research Already it seems that genomic DNA contains sequences homologous to viral enhancers[21], and what may be an enhancer has been found upstream of the sea urchin H2A gene[22] □

1 Benoist, C & Chambon, P *Nature* 290, 304 (1981)
2 Gruss, P *et al Proc natn Acad Sci U S A* 78, 943 (1981)
3 Tyndall, C *et al Nucleic Acids Res* 9, 6231 (1981)
4 Fromm, M & Berg, P Personal Communication
5 Banerji, J *et al Cell* 27, 299 (1981)
6 Moreau, P *et al Nucleic Acids Res* 9, 6047 (1981)
7 Capecchi, M *Cell* 22, 479 (1980)
8 de Villiers, J & Schaffner, W *Nucleic Acids Res* 9, 6251 (1981)
9 Sklar, J & Berg, P Personal Communication
10 Ohkubo, H & de Crombrugghe, B Personal Communication
11 Levinson, B *et al Nature* (in the press)
12 Payne, G S *et al Nature* 295, 209 (1982)
13 Katinka, M *et al Cell* 20, 393 (1980)
14 Katinka, M *et al Nature* 290, 720 (1981)
15 Fujimura, F K *et al Cell* 23, 809 (1981)
16 Sekikawa, K & Levine, A J *Proc natn Acad Sci U S A* 78, 1100 (1981)
17 Saragosti, S *et al Cell* 20, 65 (1980)
18 Jakobovits, E B *et al Nature* 285, 263 (1980)
19 Herbomel, P *et al Cell* 25, 651 (1981)
20 Jakobovits, E B *et al Proc natn Acad Sci U S A* 77, 6556 (1980)
21 Botcham, M Personal Communication
22 Grosschedl, R & Birnestiel, M L *Proc natn Acad Sci U S A* 77, 7102 (1980)

The great galactic centre mystery

from Guenter R Riegler

WHEN γ-ray astronomers discovered last year that the positron–electron annihilation radiation from the galactic centre region varied on a time scale of six months[1], it became necessary to reconcile possible explanations for the phenomenon with data from the radio and IR bands This provided the impetus for a recent workshop* that examined the central parsec of the galactic centre using data from all available spectral ranges

Images of the central few parsec of the Galaxy can be readily obtained through two experimental 'windows' bremsstrahlung emissions of ionized gas near 6 cm wavelength, and thermal emission of heated dust near 10 μm Both maps show a non-uniform brightness distribution[10] and their similarity implies that the ionized gas and the heated dust are distributed along an arc-like structure

Observations of the galactic centre at radio wavelengths show a non-thermal compact radio source surrounded by ionized gas moving at high velocities The compact source varies on time scales as short as one hour[2], but we are not yet able to form a consistent picture of the actual source size Neither ordinary pulsars nor binary stellar radio sources match the spectral index and luminosity of the compact radio source

Maps of line emission of atomic hydrogen, ionized hydrogen and molecular gas (primarily ammonia and formaldehyde) show a complex distribution at all scale sizes from a few parsec to a few kiloparsec The large, non-circular motion of a large portion of the gas may suggest it is being expelled from the centre, while the large velocity dispersion in the central parsec suggests a massive collapsed object New searches for H_2O masers in the galactic centre, reported at the workshop by Güsten (Max-Plank-Institut fur Radioastronomie, Bonn), imply that star formation is not the source of the high thermal flux in the galactic centre region

IR observations[3] show that ionized gas in the central few parsec of the galactic centre is concentrated in at least 14 small clouds An ionizing radiation of $\sim 10^{50}$ photons per s is required but its origin is debatable, sources might be distributed among, or in, each discrete IR source, or a central source of radiation might ionize all clouds

Aitken and co-workers (University College London) found that peaks in the interstellar grain temperature coincided with the positions of IR sources (as deduced from peaks in 10 5 and 12 5 μm continuum radiation), and argue that this implies that internal sources of luminosity power the discrete sources With colour temperatures derived from maps at 5, 10 and 56 μm, Rieke and Lebofsky (University of Arizona) argue that each IR source is heated by a cluster of stars of spectral type earlier than I. Gatley (UK Infra-Red Telescope) showed that most of the central parsec is transparent at UV wavelengths, so that heat sources need not be embedded in the IR-emitting dust clouds The ionizing radiation may therefore be produced by a distributed source which is merely centrally concentrated about the position of IRS 16 (commonly accepted as the IR counterpart of the non-thermal compact radio source in Sgr A West)

The main argument for the existence of a central massive object at the galactic centre comes from a study of the dynamics of ionized gas clouds Lacy[3] found that the motions of the gas clouds are best fit with a mass distribution which includes a central mass in addition to a distributed mass, each of $\sim 3 \times 10^6$ solar masses ($M_\odot$) The central mass could be a star cluster, for example, a highly evolved OB association as favoured by Lacy

The conjecture that there might be a black hole of mass 1–5 x 10^6 $M_\odot$ at the galactic centre was made more than a decade ago by Lynden-Bell and Rees[4] Arguments for and against its existence were given by several speakers at the workshop Lacy (Caltech) pointed out that a model of spherical accretion of ordinary

*A workshop on the galactic centre was held on 7–8 January 1982 at the California Institute of Technology The proceedings will be available from the American Institute of Physics, New York, in June 1982

Guenter R Riegler is in the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109

0028-0836/82/180018-02$01 00

interstellar gas towards a massive black hole would be inconsistent with the observed Sgr A IR spectrum and the required ionizing flux The accretion rate which is required to power all activity in the galactic centre region, $\sim 10^{-5}$ $M_{\odot}$ yr^{-1}, may also be augmented by debris from stellar disruptions or from nearby collisions involving giants Some authors, particularly Ozernoi and his collaborators[5], have argued against the massive black hole hypothesis on the grounds that the rate of star capture would be incompatible with observations Rees (University of Cambridge) reviewed these arguments and showed that it is nevertheless possible to construct models in which the luminous flare from each stellar remnant's impact with the black hole would not lead to excessive average luminosities

Charge-coupled device images of the galactic centre at an effective wavelength of 0 9 μm came from three groups and all showed a pair of faint, very red sources within a few arc seconds of IRS 16 and the compact non-thermal radio source The sources are not as highly reddened as had been expected and it was suggested that line emission from doubly ionized sulphur may contribute to the observed photon flux within the bandpass of the CCD/filter combinations

The positron–electron annihilation line at 511 keV has been observed by at least five experimental groups over the last decade and is, in fact, the only non-solar γ-ray line that has been confirmed by multiple measurements The line emission[6] implied an annihilation rate of 10^{43} per s, and a luminosity of 10^{37} erg s^{-1}, compared with an observed luminosity at IR wavelengths of 10^{40} erg s^{-1} (the bolometric luminosity of the source of ionizing radiation in the galactic centre must be still higher by roughly an order of magnitude[3]) Furthermore, the luminosity of the annihilation radiation exceeds the radio luminosity (1 35–31 cm) of the non-thermal source[7], and is a few hundred times more intense than the X-ray luminosity, measured by the Einstein Observatory from the central few parsecs of the Galaxy[8]

The first satellite measurements of the 511 keV line from the galactic centre in 1979–80 showed that the emission had decreased by a factor of roughly three over a time span of six months[1] Recent experiments by Leventhal (Bell Laboratories) and Paciesas (Goddard Space Flight Center) did not observe a positive signal for line emission from the galactic centre, consistent with the assumed further decrease in intensity Jacobson (Jet Propulsion Laboratory) pointed out that the available data are statistically consistent with a model consisting of an extended, constant-intensity source and a point-like, variable source of 511 keV line emission

If the variable positron emission comes from a point source at the galactic centre, then some inferences can be made about the positron source as well as the positron annihilation region As pointed out by Ramaty (Goddard Space Flight Center) and Lingenfelter (University of California, San Diego), the width of the 511 keV line and its temporal variation[1,6] require that the annihilation region have a temperature, density, degree of ionization and size which are consistent with the peculiar warm clouds[3] and the other compact IR sources[7] observed in the central parsec of the Galaxy

The nature of the positron source is also strongly constrained by the observed line emission variation multiple, extended sources such as pulsars, supernovae, black holes, or cosmic-ray interactions with the interstellar medium are excluded Blandford (Caltech) suggested that the positrons might be produced by an electromagnetic cascade process in the magnetosphere of a $\sim 10^6$ $M_{\odot}$ black hole In this model, high-energy ($\gg m_e c^2$) photons and electrons interact with much cooler ambient photons and gas Lingenfelter and Ramaty point out that interactions in which the photon, electron and ion energies are all comparable, and of the order of $\sim m_e c^2$, are much more efficient for electron–positron pair production The necessary energy could be provided either in the vicinity of a very compact source, like a black hole of $< 10^2$ $M_{\odot}$ (to agree with arguments by Ozernoy that the tidal disruption of stars would otherwise be inconsistent with stellar luminosity in the vicinity of the galactic centre) or by interactions of high-energy beamed photons with one another

Oort (Leiden Observatory) described the galactic gas motion of scale sizes of ~ 1 pc, ~ 300 pc and 1–3 kpc The non-circular motions and the tilted distributions of the molecular clouds in the two larger-size regimes have been ascribed to either eruptive activity from the nucleus, or large-scale streamings in a non-axisymmetric potential field, but a unified model is apparently not yet available Recent studies by Lake (Bell Laboratories) of orbits in three-axial ellipsoids have shown the existence of 'anomalous' orbits in tilted planes which may explain at least part of the observed features on the 1–3 kpc size scale The molecular clouds within 300 pc are, however, more likely to originate from an eruption in the nucleus of the Galaxy □

1 Riegler, G R *et al Astrophys J Lett* **248**, L13 (1981)
2 Lo, K Y , Backer, D C & Cohen, M H (in preparation)
3 Lacy, J H *et al Astrophys J Lett* **227**, L17 (1979)
4 Lynden-Bell, D & Rees, M J *Mon Not R astr Soc* **152**, 461 (1971)
5 Gurzadyan, V G & Ozernoi, L M *Astr Astrophys* **95**, 39 (1981) and refs therein
6 Leventhal, M *et al Astrophys J Lett* **240**, L338 (1980)
7 Lo, K' Y *et al Astrophys J* **249**, 504 (1981)
8 Watson, M G *et al Astrophys J* **250**, 142 (1981)
9 Brown, D L *et al Astrophys J* **250**, 155 (1981)

Are Sahelian droughts predictable?

from S I Rasool

WHILE the controversy over whether the Earth as a whole is cooling down or warming up continues to rage unabated, there are a few large-scale climatic phenomena with major economic and social repercussions that are becoming amenable to analysis One is the Sahel drought problem Others include the circumstances which lead to the onset of the Indian monsoon and the impact of Pacific sea-surface temperature anomalies on North American climate and the upwelling of cold water off South America, with its important effects on fisheries and global climate Unlike these latter phenomena, however, the Sahel drought has not been the object of coordinated measurements and intensive research The lack of activity is all the more conspicuous in light of the enormous increase in the amount of climatic data acquired over Sahel in the last ten years

Landsats 1, 2 and 3 have now provided about ten years of uninterrupted observations over Sahel which can be used to estimate temporal changes in the surface cover and in the albedo The man-made increase of albedo, impeding convection, has been put forward as the principal factor triggering the drought The energy balance over Sahel and its variation at different times in 1978 and 1979 has also been successfully measured with Meteosat I Meteosat II has now become operational and will continue to provide information on cloud cover, and on ground albedo, surface temperatures and atmospheric water vapour over Africa and the Atlantic NOAA meteorological satellites are also giving daily coverage of meteorological parameters over most of the Earth But how can one get the people with this vast quantity of data talking to those who need the data?

COSPAR (International Committee on Space Research) is keen to see how observations from space can help resolve some of the fundamental issues in the earth sciences and decided to sponsor a workshop on the Sahel climate problem The OECD's Club du Sahel, which is interested in the economic development of

S I Rasool holds a Fondation de France Chair of Atmospheric Science at the Laboratoire de Météorologie Dynamique, Ecole Normale Superieure, Paris 75231

0028-0835/82/180019-02$01 00

the Sahelian countries, hosted the workshop and it was held in Paris in January

The Sahel drought of 1969–73 was not considered to be unique as there have been other long periods of severe drought, notably in the 1910s and 1940s Detailed analysis of data from some 1,200 stations in Africa (S Nicholson, Clark University) suggests that the Sahel drought is related to rainfall anomalies on a continental scale, extending from southern Africa to just north of the Sahara desert, which characteristically persist for as long as 10 to 15 years at a time Among possible explanations for recurrent long-term climate anomalies are the displacement of the Inter-Tropical Convergence Zone (for yet to be discovered reasons) and changes in global east to west air circulation patterns caused, in part, by changing conditions over distant oceans (Indian, Atlantic and Pacific) leading to lack of moisture over the Sahel at the beginning of the rainy season

The persistence of drought may in part be caused by biogeophysical feedback mechanisms involving long-term changes in the heat balance of the Sahel, partly driven by changes in surface properties (such as albedo and soil moisture) Modelling studies (P Rowntree, British Meteorological Office) indicate that soil moisture in the Sahel may have a significant effect on the circulation of the atmosphere and on rainfall over West Africa Analysis of Landsat data (M F Courel, IBM Scientific Center, Paris) indicates an increase in surface cover and a decrease in surface albedo over northern Senegal in the past eight years

The amount of dust blown from the Sahara Desert over the Atlantic Ocean is enormous, ranging from 100 to 400 million tons a year, and making up as much as one-half of all the dust in the Earth's atmosphere The dust may change the energy balance of the ocean–atmosphere system over the Atlantic considerably (T Carlson, Pennsylvania State University) and affect the climate of the Sahel and possibly of the whole world, although the precise magnitude of the phenomenon is not yet known

Scientists from COSPAR at the meeting recommended that intensive analysis be made of surface and radiosonde data from the Sahel and Central Africa for the past 15 years Such data are available but have not yet been completely assembled and analysed Coordinated and intercalibrated analysis of satellite data from pre-drought, maximal drought and post-drought periods would give a better understanding of the mechanisms behind the onset of the drought, and its persistence in 1969–73 Over the past 10 to 15 years, satellites have furnished a large body of observations of the Sahel, neighbouring regions of Africa and the Atlantic from which information on cloud cover, dust, surface temperature, surface albedo, heat balance and soil moisture could be extracted The COSPAR scientists also suggested that a field experiment in the Sahel, within a 200×200 km area, be carried out to measure the vertical and horizontal fluxes of moisture and heat Such a field experiment, within the grid of a standard general circulation model, could be designed based on existing ground and satellite networks and would provide a critical test for the general circulation model.

The scientists attending the meeting urged COSPAR, in cooperation with the World Climate Program and the Club du Sahel of the OECD, to create an *ad hoc* working group to speed up implementation of these recommendations All the necessary ingredients seem to be there an enigmatic problem of extreme economic and social importance, the existence of large amounts of unanalysed data both from ground and space, and good scientists apparently available to start working on the problem Get the right data to the right people and we may actually begin to see what causes the droughts in Sahel □

Meteoritics: evidence for chemical fractionation in the early Solar System

from Robert Hutchison

CHONDRITES, the most numerous of meteorites, are stony types which have not been melted since they formed some 4 55 Gyr ago They are complex and contain not only minerals that formed at high temperatures, but also volatiles that seem unlikely to have condensed from a gas of solar composition unless the temperature was below 700K The siting of these volatiles has a bearing on theories of the formation of the ordinary chondrites which, in turn, place important constraints on theories of planetary formation There are three main theories, explained below, of which the first two have the most supporters

(1) When the various materials accreted, different proportions of volatile-rich fraction were included Different parts of chondritic parent-bodies were subsequently heated and metamorphosed to different degrees, but without loss of volatiles An unequilibrated chondrite, such as Tieschitz, thus contained, on formation, a higher proportion of 'Holy Smoke' (ultrafine-grained, volatile-rich matrix) than a metamorphosed member of the same chemical group, such as Allegan[1]

(2) Each chemical group of the ordinary chondrites had, on accretion, a uniform composition Heating and metamorphism occurred in conditions where volatiles were lost, the loss being greatest in the types subjected to the highest temperature[2]

(3) A theory favoured by a minority, including myself and some of my co-workers, is that ordinary chondrites accreted when hot (1,070± 100K) Volatiles were most readily introduced into the material nearest the surface, which cooled most rapidly[3–5] No post-accretionary heat source is required

It is clear that the first and third hypotheses lead to a theory of heterogeneous accretion of planetary bodies, while the second favours homogeneous accretion Choice of the most plausible theory is obviously crucial to our understanding of the origin of the Earth and makes identification and separation of the volatile-rich fraction in chondrites of particular significance

It is not surprising, then, that over the past two decades there have been many mineralogical and chemical investigations of chondrites aimed at deciphering their thermal and chemical histories in the hope that part of the story of planetary formation may be pieced together

Somewhat contradictory conclusions were reached in three of the most recent contributions to this field Rambaldi *et al*[6] argued that in ordinary chondrites, the volatiles tend to be concentrated in 'Holy Smoke', some of which may even be identifiable in chondrite types which equilibrated at high temperature In contrast, Ashworth[7,8] identified ultrafine-grained, 'non-clastic' matrix as a minor component in a few highly unequilibrated ordinary chondrites only Finally, Christophe-Michel-Lévy[9] suggested that equilibrated ordinary chondrites accreted over a range of temperatures extending up to 1,270K, whereas an unequilibrated H-group chondrite accreted below 620K, only in this last case was fine, volatile-rich dust incorporated

Part of the problem of these conflicting interpretations lies in the strengths and weaknesses of the different techniques used Rambaldi *et al*[6] took samples of six ordinary chondrites and used freeze–thaw and/or gentle disaggregation followed by sieving to produce various-size fractions down to 0 08 μm Instrumental neutron activation analysis was used to determine up to 24 elements in each size-fraction and further, as yet unpublished, information, has been derived from the electron microprobe and the scanning electron microscope (SEM)

Ashworth used the high-voltage electron

Robert Hutchison is Curator of the Meteorite Collection at the British Museum (Natural History), Cromwell Road, London SW7 5BD

0028-0836/82/180020-02$01 00

microscope on ion beam-thinned samples to identify minerals and examine textures at the submicrometre level, but detailed work is of necessity limited to very small areas Christophe-Michel-Lévy's SEM approach looked at textures and minerals, with semi-quantitative chemical information on the latter, in essentially untreated fragments of meteorite Each of the three approaches provides useful information which cannot be obtained using the other techniques, the chemical data of Rambaldi *et al.* cannot yet be had from single, sub-micrometre-sized grains

Rambaldi *et al* [6] found that the finest size-fraction of each meteorite — that richest in 'Holy Smoke' — was generally richest in moderately volatile and volatile elements such as K, Zn and Sb, these fractions tended also to be enriched in refractory elements such as W and the rare earths This is true of the H5, equilibrated chondrite, Allegan, but, except for Zn, Ag and W, not true of the unequilibrated, H3 Tieschitz And here lies the problem Although the distribution of the 24 elements between the size-fractions may be explained in general terms, difficulties arise with individual meteorites

For example, Co, Ni and Fe are enriched in the finest fraction of Tieschitz relative to the two coarser fractions Rambaldi *et al* [6] suggest that this may be due to the presence of a Ni-rich magnetite However, Christophe-Michel-Levy[9] found the finest-grained matrix to be mainly iron-rich olivine, and Ashworth[8] states that this 'dark matrix' has "troilitic and metal-rich varieties" Neither author found magnetite in Tieschitz It may be concluded, therefore, that Rambaldi *et al* 's finest size-fraction probably contains a mixture rich in olivine, metal and troilite, but with other phases too A further observation of Christophe-Michel-Lévy[9] helps clarify the reason why the rare earths La and Sm are depleted in the finest size-fraction of Tieschitz She noted that in this meteorite, phosphate is associated with Fe–Ni metal, occurring largely as inclusions within metal grains Phosphate has also been found as tiny crystals in cracks in metal where it probably formed from P, expelled from metal, by reaction with O and Ca from neighbouring 'dark matrix' (A W R Bevan and R Hutchison, unpublished) Phosphates concentrate rare earths, and, although the crystals are fine-grained, most (and most rare earths) probably would have been retained in metal during the gentle disaggregation carried out by Rambaldi *et al* [6]

So what of 'Holy Smoke'? Search for a carrier of volatiles in the ordinary chondrites essentially began in 1973. In that year, Anders' group discovered that some whole meteorite samples of the brecciated H-group chondrite, Supuhee, are enriched in Tl and Bi relative to CI chondrite[10] For several reasons CI chondrites are taken to be our best available sample of unfractionated condensable Solar System material Enrichment of the volatile metals Tl and Bi in bulk Supuhee was therefore interpreted as indicative that a fraction must exist which is enriched in these metals by some hundreds of times relative to CI This is because Supuhee is composed mainly of high-temperature minerals and so the volatile-rich fraction — termed 'mysterite' — cannot occupy more than a few per cent by volume Anders' group concluded that 'mysterite' represents a rare substrate on which Tl and Bi condensed at a late stage in the cooling of a solar nebula

'Mysterite' or 'Holy Smoke' must, however, be widely distributed among ordinary chondrites. Tieschitz, for example, contains more Bi and as much In and Tl as CI[10] Dodd[11] and others have shown that this stone contains only 8 per cent by volume of material with a grain size less than 100 μm, and it has already been stated[8,9] that in this fraction most grains are of high-temperature minerals If water and carbon, with other volatiles, are concentrated in the remainder of the fine-grained matrix, then this can constitute no more than about one per cent of the whole meteorite by volume It is the search for, and characterization of, this type of material that now looms large in meteorite research[12] □

1 Takahashi, H , Gros, J , Higuchi, H , Morgan, J W & Anders, E *Geochim cosmochim Acta* **42**, 1859 (1978)
2 Wasson, J T *Meteorites* (Springer, Berlin, 1974)
3 Hutchison, R , Bevan, A W R , Agrell, S O & Ashworth, J R *Nature* **280**, 116 (1979)
4 Hutchison, R , Bevan, A W R , Agrell, S O & Ashworth, J R *Nature* **287**, 787 (1980)
5 Bevan, A W R & Axon, H J *Earth planet Sci Lett* **47**, 353 (1980)
6 Rambaldi, E R , Fredriksson, B J & Fredriksson, K *Earth planet Sci Lett* **56**, 107 (1981)
7 Ashworth, J R *Earth planet Sci Lett* **35**, 25 (1977)
8 Ashworth, J R *Proc R Soc* **A374**, 179 (1981)
9 Christophe Michel-Levy, M *Earth planet Sci Lett* **54**, 67 (1981)
10 Laul, J C , Ganapathy, R , Anders, E & Morgan, J W *Geochim cosmochim Acta* **36**, 329 (1973)
11 Dodd, R T *Earth planet Sci Lett* **30**, 281 (1976)
12 Higuchi, H , Ganapathy, R , Morgan, J W & Anders, E *Geochim cosmochim Acta* **41**, 843 (1977) and references therein

The aging of the brain

from Patrick Rabbitt

CONFRONTING an immanent future in which one-fifth of the population of the developed countries will be aged over 65, delegates to a conference* on the aging brain need not lack large general questions to attack It is a comment on the present underfinanced and fragmentary state of research on aging that most papers were disjunctive presentations by research teams primarily known for their expertise in other areas None of these teams can yet afford to commit all their resources to the systematic study of the aging central nervous system (CNS) For this reason there could be little discussion between groups interested in psychology and behaviour, in neuroanatomy and physiology, in neuroendocrinology, in neurochemistry and pharmacology or in animal models of aging

Several important general problems of the aging CNS were illustrated, rather than illuminated, by the context in which they were raised D J Reis (New York) and his colleagues neatly pointed out the increasing variability in performance which occurs among aging animals This can be due to different rates of cell loss, both between different animals and between different areas of the same brain Alternatively, rates of loss may be constant across different brains and brain areas, but this loss may start from different initial levels of neuronal density Or, of course, variability in performance may be due to interactions between both these factors Reiss and his group found differences in local brain weight in the caudate nucleus and the olfactory tuberule in BALB and CBA mouse strains These differences correlated with differences in behaviours thought to be mediated by dopaminergic neurone systems in open fields and in habituation to exploration of objects in dark slots in a test box The strains also show different responses to amphetamine This raises the fascinating possibility that specific differences in particular parts of the brain related to the presence or absence of single genes, or groups of genes, provide a direct link between genes, brain anatomy, neurochemistry and behaviour The implications for relative cell loss with aging seem secondary to this important news

For functioning human beings the most intriguing question about the aging of the brain is whether, individually or universally, we suffer a steadily accelerating loss of ability after some brief peak in our early twenties or whether, more happily, we can look forward to a long plateau of optimal function followed by a mercifully sudden 'terminal drop' in performance just before death which makes it irrelevant whether we have our wits about us or not L F Jarvik (University of California, Los Angeles), who has studied a particular group of people, including some monozygotic twins, since the late 1940s, is one of the few investigators in the world who has data bearing on this question She discussed

*The 10th Aharon Katzir-Katchalsky Conference on 'The Aging of the Brain' was held in Mantua, Italy on 26–29 March 1982

Patrick Rabbitt is Professor in the Department of Psychology, University of Durham, Durham DH1 3LE

0028-0835/82/180021-02$01 00 R423 4013

psychometric data on 35 of these people as their average age rose from 64 to 73 years and found, overall, no significant decline and, indeed, an apparent improvement on some sub-tests as age increased From 73 to 84 years the picture was less encouraging and, when group mean scores were considered, all test scores declined

This does not yet answer the question since group mean scores might decline because particular individuals within the group might become ill and approach death just before testing, and the proportion of people showing such terminal changes in performance and contributing to mean scores would be certain to increase as the group aged Jarvik's analyses of her data bore out this possibility Individual subjects showed little or no decline in performance through the seventies and into the eighties Moreover, abrupt declines in performance in individuals were good statistical predictors of impending death Equally interestingly, elderly peoples' self-rating of their own health status was also a good predictor of their actual future survival — as good as, but (perhaps disappointingly!) no better than, their physicians' predictions

All longitudinal studies are certain to be criticized because at the time when they yield their richest data the focus of scientific interest is bound to have changed One cannot, alas, time travel 20 or 30 years to include cases or use tests which become crucial as the state of the art advances However, investment in really adequate longitudinal studies need only be very modest Such studies are expensive to initiate, but very cheap to run once they are begun Jarvik's study illustrates, but does not resolve, the great simple questions about changes in human efficiency with advancing age If we want to know more about our own personal futures, and about the futures of the societies we have to plan, no single investment can yield so much useful information for so modest an outlay □

Engineering organic molecular layers

from C W Pitt and I A Shanks

Scientists and engineers are taking a renewed interest in thin films of organic materials deposited, molecular layer by molecular layer, by an unconventional process attributed to Irving Langmuir and Kathleen Blodgett and dating from the 1930s Such films can exhibit a remarkable perfection of structure over large areas and as the successive monomolecular layers may be of the same or different materials, a kind of 'molecular engineering' is made possible

Long hydrocarbon chain molecules, with a hydrophilic group at one end and a hydrophobic group at the other, can be made to form a continuous film by spreading them on a water surface The nature of the groups at the ends of the molecules ensures that one tip of the molecule is immersed in the water while the other remains in the air A movable barrier which penetrates the water surface may be used to compress the film into a condensed layer exactly one molecule thick in which the oriented molecules are held together by van der Waals' forces The film may then be transferred from the water surface on to a solid substrate by dipping the substrate through the surface and then withdrawing it whilst continuing to compress the film Repeating this process allows a multilayer film to be progressively built up

The Langmuir-Blodgett (L-B) technique undoubtedly emerged at a time long before the technology to exploit its unique capabilities was available A recent three day symposium* made it clear that the situation has now very much changed Recent work has highlighted materials that have attributes which can be readily exploited when they are deposited as L-B layers Many of these material systems have been developed as modified molecular structures tailored specifically for this deposition technique.

Typical of the molecular structures reported were the diacetylenes — a monomer/polymer system incorporating a conjugated triple bond which may be induced to cross-link using an electron beam or by UV or visible radiation Diacetylene films are being investigated as gate region insulators in indium phosphide field effect transistors and as layers in optical waveguiding devices Double bond molecules, such as ω-ticosenic acid, have been deposited as very thin electron beam resists capable of defining lines less than 600Å wide, while even the simple fatty acids have found application, in very pure form, in modifying the surface properties of semiconductors used in Schottky barrier diodes and photodiodes

A radically new approach to modelling biological membranes is also offered by the L-B technique The fluid bilayer of lipid molecules may be simulated by a simpler structure fabricated from reconstituted purified natural membrane components and laid down by the L-B process Protein complexes, such as bacteriorhodopsin, can be immobilized in such layers, simulating the natural system Possible applications of these biocomposite films include filtering solutions with chemically selective elements incorporated in the L-B layer Entirely synthetic membranes, such as polypeptide layers, have also been fabricated, although applications are less apparent at this stage

A technique for retaining the interesting properties of non-amphiphilic molecules, while enabling L-B deposition, is clearly emerging, that of attaching a hydrocarbon chain and a polar group to an aromatic ring-type molecule Tetraphenylporphyrins, anthracene, perylene and pyrene were all reported as molecules which have been treated in this manner Applications include tunnelling layers and electro-fluorescent display panels

In addition to using these films in device applications they are also of interest as investigative tools in the understanding of a wide range of other phenomena Examples are the use of manganese stearate monolayers to explore existing theories of two-dimensional magnetism and of monolayers of fatty acids and their salts to probe the mechanisms of the sliding friction between surfaces The similarity between L-B multilayers and liquid crystals has also been noted and papers were given on the deposition and properties of *n*-pentyl cyanoterphenyl and other liquid crystalline materials □

*A three day symposium of forty invited scientists and engineers was organized by the Rank Prize Funds to explore the science and optoelectronic applications of Langmuir- Blodgett layers The participants included both junior researchers in the field and eight invited speakers from Europe, the US and the UK

C W Pitt is in the Department of Electronics and Electrical Engineering, University College London WC1E 6BT, and I A Shanks at the Royal Signals and Radar Establishments, Malvern, Worcs

100 years ago

Rarely has so distinguished and representative an assembly been seen in Westminster Abbey as that which met to pay the last honours to Mr Darwin, on Wednesday last week The Abbey indeed was crowded The character of the long line of distinguished men who followed the honoured remains to the grave, may be seen from the list of pall-bearers — The Duke of Devonshire, the Duke of Argyll, the Earl of Derby, Mr J Russell Lowell, the American Minister, Dr W Spottiswoode, P R S , Sir Joseph Hooker, Mr A R Wallace, Prof Huxley, Sir John Lubbock, and the Rev Canon Farrar, Mr Darwin has been buried close beside the grave of Sir John Herschel, and within two paces of that of Sir Isaac Newton
From *Nature* **26**, 16, 4 May 1882

0028-0836/82/180022-01$01 00

REVIEW ARTICLE

Human body clocks and the timing of sleep

Arthur T. Winfree
Department of Biological Science, Purdue University, and Institute of Natural Philosophy, West Lafayette, Indiana 47906, USA

The alternation of sleep and wakefulness in humans exhibits surprising regularities during long-term isolation from the usual time-of-day cues. The inferred domination of our consciousness and physiological states by internal clocks may soon have practical implications for medicine and psychiatry.

WE ALL fall asleep from time to time and no one seems to know why[1-4] While obliged to adhere to a daily schedule of work and rest, light and darkness, most people go to sleep and wake again once a day at predictable times The regularity of this schedule is imposed ultimately by the rotation of the whole Earth, but it turns out that rhythmicity persists even in the absence of any imposed schedule[5-11] There are some intriguing regularities about the timing of sleep onset and the subsequent wake-up in these conditions of so-called 'temporal isolation' They reveal the presence of an internal clock with natural period just a little longer than the Earth's rotation period Aschoff and co-workers were the first to recognize the influence of this clock on the timing of sleep and in the daily rise and fall of body temperature[5 6 11 12,13] There may even be two timers one is a circadian clock with spontaneous period 24–25 h dominating the rhythms of body temperature, plasma cortisol and rapid eye movement sleep, the other seems to have a longer period governing the ebb and flow of sleep and associated physiological rhythms such as growth hormone secretion[14-20]

Prolonged temporal isolation

Inferences about the number and variety of timers affecting our sleep habits are drawn from data similar to those of Fig 1 Here we see the times of sleep (black) and waking (white) of an individual living without external time reference for 4 months in a cave[21 22] Figure 1 replots the original data in the TV-raster format now conventional the time axis is broken up into consecutive equal intervals (in this case, of 24 3 h) stacked from top to bottom The whole picture is then duplicated to the right to emphasize the continuity of time from the right edge of each 24 3-h interval to the left edge of the next

The black bars would be stacked in a vertical column if sleep and waking recurred at exact 24 3-h intervals They do not The intervals intermittently grow longer, shearing the stack of black bars to the right as it descends The raster plot in Fig 1 conspicuously depicts a slow but relentless change of organization in the timing of sleepiness Czeisler[23] was the first to notice this in a diversity of such records It appears that briefer records can often be assigned a place in a similar progression sleep onsets tend to recur at longer and longer intervals, while still preferentially occurring near the temperature minima (recurring at 25-h intervals) Near the bottom of Fig 1, sleep and waking alternate about once in every two cycles of 24 3 h. This phenomenon has been observed in humans at least since the cave experiments of Chouvet *et al*[21,22] Isolated subjects, emerging from a month's studious or meditative confinement, exhibit understandable perplexity, even disbelief, that the time has passed twice as fast as they reckoned. During these 2 1 intervals, one subject on average slept 19 h at a stretch and stayed awake 31 hours before retiring again, never aware that his days were abnormally long Such individuals typically eat somwhat more at three meals between rising and retiring, but not enough more to prevent weight loss (E D Weitzman, personal communication) Needless to say, this was at first alarming to observers outside, responsible for the subjects' well-being (ref 11, p 59)

Whatever the reason for this gradual drift, it seems to have little effect on another regularity independently noted several years ago by Zulley[24-28] and Czeisler[23] Figure 2 shows the durations of many sleeps in three independent experiments in conditions of temporal isolation Sleep duration, plotted vertically, is arranged horizontally according to the timing of sleep onset relative to the temperature rhythm In Fig 2*a*, sleep onset is measured from the previous temperature minimum In Fig 2*b* it is measured from a regression line through a long series of temperature minima found to recur on average every 24 9 h In Fig 2*c*, using the data of Fig 1, it is measured from a regression line at a 24 33-h period, indirectly inferred to parallel the temperature rhythm In all cases sleep durations are shorter for sleeps undertaken later—but only up to a point sleeps undertaken too late after the temperature minimum abruptly become long again

Our usual 8-h sleeps fall near the middle of this ramp There is some question about whether the smooth curve one might draw through the descending cloud of dots also rises to connect the cloud's lower end to its upper end this gap may be a discontinuity typical of rhythmically modulated threshold mechanisms[29] There being few sleeps of middling duration (if any) that begin near this critical time, a gap devoid of wake-ups appears in the raster plot (shaded in Fig 1) There is no corresponding gap devoid of sleep onsets, nor do sleep durations fall into any clear pattern when plotted against sleep onset time, nor can sleep onset time be anticipated with precision comparable to the anticipation of wake-up timing implicit in Fig 2 Sleep onset and wake onset seem to be governed by different principles

Mathematical allegories

A variety of mathematical models have been postulated to describe the regularities noted above[11 13 18 29-34] For the present they remain more nearly computational metaphors than testable deductions from physiological hypothesis However, the more sophisticated among them (see, for example, refs 18, 30) may soon prove useful in disentangling the relative contributions of various possible cues to entrainment of human body clocks Without such cues, and without their somehow perturb-

0028-0836/82/180023—05$01 00

ing the mechanisms of our internal clocks, we could not keep our habits synchronous with the day/night cycle. Such cues may include mere knowledge of time of day[8,9 13 35], light/dark transitions (ref 36 p 20, and refs 32, 37–40), compulsory sleep itself[11,33 36 37 40–44], food and stimulants[44–48], or even a young man's delight to receive illicit daily letters from a young woman lab assistant while the professor was out of town (ref 11 p 36, and R Wever, personal communication)

The supra-chiasmatic nuclei

The physiological sources of circadian timing seem to lurk somewhere in the central nervous system[49–54] Control of sleepiness/activity timing in mammals probably resides in a pair of tiny structures called the suprachiasmatic nuclei (SCN) in the hypothalamus, astride the crossing of the optic nerves[55–59] Only last year was an anatomical locus presented as the SCN in man[60] Single cells in this region respond to visible light in rough proportion to the log of light intensity striking the eyes[61–64] (The primary visual areas, in contrast, respond mainly to spatial and temporal gradients of light intensity) This absolute intensity information is conveyed along the retinohypothalamic tracts, identified in a wide variety of mammals (including primates but not yet in man) They start in the left and right retinal ganglion cells and ramify (with mixing) in the left and right SCN[65] Stimulation of this pathway by light or by electrical impulses does alter the firing of SCN neurones[61–63,66] The aggregate firing of SCN cells increases and decreases 10-fold with circadian period in ostensibly constant conditions (ref 67 and S T Inouye and H Kawamura, personal communication) In rats the metabolic rate of SCN tissue, assayed by 2-deoxyglucose uptake also continues rhythmically in constant conditions[68–70] An island of SCN isolated from its neural surroundings in the rat still shows a persistent circadian rhythm of neural activity, while similar rhythmicity no longer occurs elsewhere in the brain[67] It follows that if there is no biochemical or thermal rhythm impinging on the neurally excommunicated SCN, then the SCN in this species is indeed a self-contained oscillator

Combining observations on rats and monkeys, we might conclude that sleep/wake timing is indeed governed by at least two independently competent clocks But no one yet knows whether both function in any one species, nor in particular, whether they do in man

Artificial light: too dim to cue human body clocks?

Our internal clocks normally remain quite predictably synchronized to the 24-h period of environmental cues (and therefore also to each other) The question of what constitutes a compelling cue is obviously of vital importance and is the subject of vigorous experimental investigations The anwers differ significantly from one species to the next Bright light may be a universal cue, its synchronizing role has been doubted only in the case of man

Activity in the SCN suppresses release of melatonin from the pineal gland, so does visible light, presumably acting through the SCN A melatonin secretion rhythm seems to free-run in synchrony with core temperature (A J Lewy *et al*,

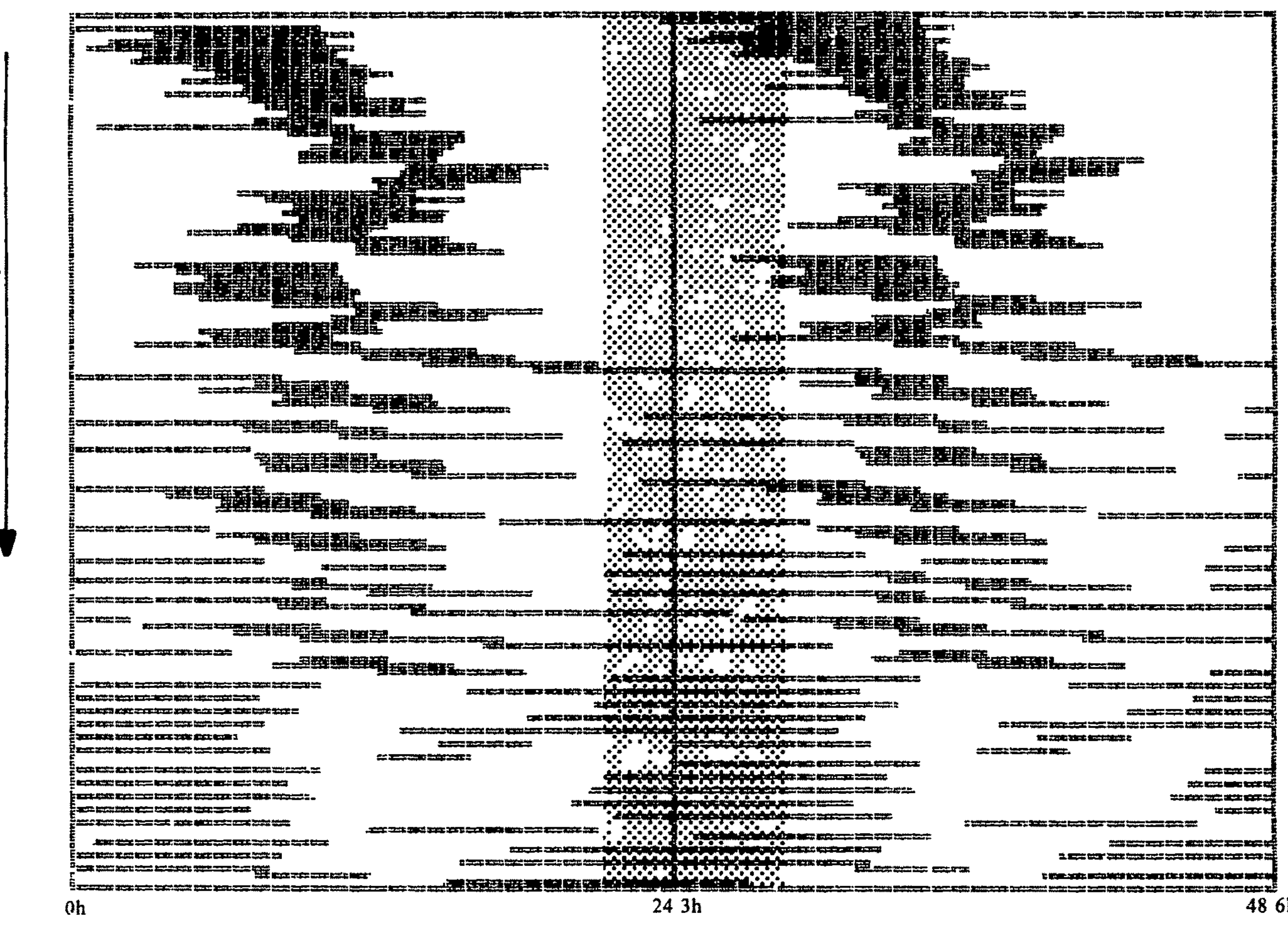


Fig. 1 Data[21,22] of digitized and reassembled double-plotted raster format following Czeisler[23] Sleep time is indicated by a black line from left to right, with successive 24 3-h intervals stacked from top to bottom This diagram is really a cylinder of 24 3-h girth along which alternating intervals of sleep and waking are plotted along a helically wound time axis It is slit open and laid flat for the publisher's convenience, the whole picture is replicated to the right, shifted up a line, to exhibit its continuity across the arbitrary slit line Notice the shaded range of phase in the 24 3-h cycle when spontaneous wake-up seldom occurs There is no corresponding range devoid of sleep onsets

in preparation) The immediate and immediately reversible effect of light on melatonin release thus provides one assay (as yet uncalibrated) for the impact of light on the sleepiness/activity timer

Quantification of this impact has been sorely needed since the discovery[13] that in certain conditions light may be a less effective synchronizer than social stimuli (in man)[11 36,37] Perlow *et al*[71] showed that room light inhibits melatonin secretion in primates Do the several studies showing lack of comparable impact in humans reveal a basic neuro-anatomical or physiological change unique to humans? Richter[72] presents these speculative matters from a rather different viewpoint, that people override their internal clocks with the aid of artificial light, supposing that artificial light is as good as sunlight for these purposes

Lewy *et al*[73 74] provide a first quantitative estimate of the impact of light in man by monitoring plasma melatonin during exposure to light of various intensities The results suggest instead that humans are not altogether insensitive but rather only much less sensitive to dim light than are the primates and nocturnal rodents examined to date In particular, intensities (about 500 lux) comparable to those of indoor artificial illuminants used in experiments throughout the past decade have little or no effect on melatonin secretion The light/dark alternations used in those experiments may have been dark/dark so far as the sleepiness timer is concerned

Intensities of about 2,500 lux (more like outdoor sunlight) are effective in man If similar exposures are needed to phase-shift the circadian rhythm, it could clarify much of the debate over man's entrainment more by social stimuli than by the light/dark cycles that cue every other organism It also suggests an understanding of Klein and Wegmann's observation[75] that the symptoms of 6 hours' jet lag persist longer in transatlantic passengers who retire to their hotel rooms to sleep it off, than in others imposed upon to go outdoors[73]

Therapy for mood and sleep disorders

In some individuals melatonin secretion is suppressed by dim light Lewy *et al*[73] report suppression by as little as 500 lux incandescent light in four subjects who are peculiarly susceptible to manic-depressive alternations This observation enhances interest in recent conjecture[74 76–80] that fluctuations of mood, including manic-depressive alternations, may be caused by abnormal mutual phasing of circadian rhythms Wehr *et al*[78] have found that by deliberately advancing the sleep/wake rhythm by 6 h relative to the temperature rhythm, they could sometimes abort depressive episodes in one patient More recently in another patient, Lewy *et al*[81] precociously terminated recurrent winter-time depression by artificially lengthening the photoperiod

Phase relations between sleep and core temperature can also be quite different in individuals who for some reason are not normally responsive to the usual synchronizing cues[82] Among sightless individuals, those who do entrain show diverse phase angles relative to the civil and solar day[83 84] Some sightless individuals, some recluses and older people living indoors with little social contact, and perhaps individuals with stronger circadian clocks, sometimes retire and rise later and later every day like the tides Eventually, pursuing their solitary interests by night and sleeping by day, 180 degrees out of phase with surrounding society, they drift still later into synchrony again after another 2 weeks or so[85–88] (see Fig 3 for an example)

All subjects fail to synchronize to scheduling at too short or too long a period relative to their own internal clocks. Wever[89] has exposed subjects in his temporal isolation facility to gradually lengthening (or shortening) days As their activity cycles shift with respect to the light/dark cycle and their temperature cycles shift with respect to their activity cycles, they gradually become 'early birds' or 'night owls' This becomes noticeable subjectively only when their usual timing of waking or retiring transgresses a boundary of daylight However, when the driving period becomes too extreme (beyond 22 or 27 h), both internal and external synchronization fail: neither core temperature fluctuations nor sleep/wake alternations follow the light/dark cycle, nor do they remain entrained to each other Curiously, the subject remains unaware of any change

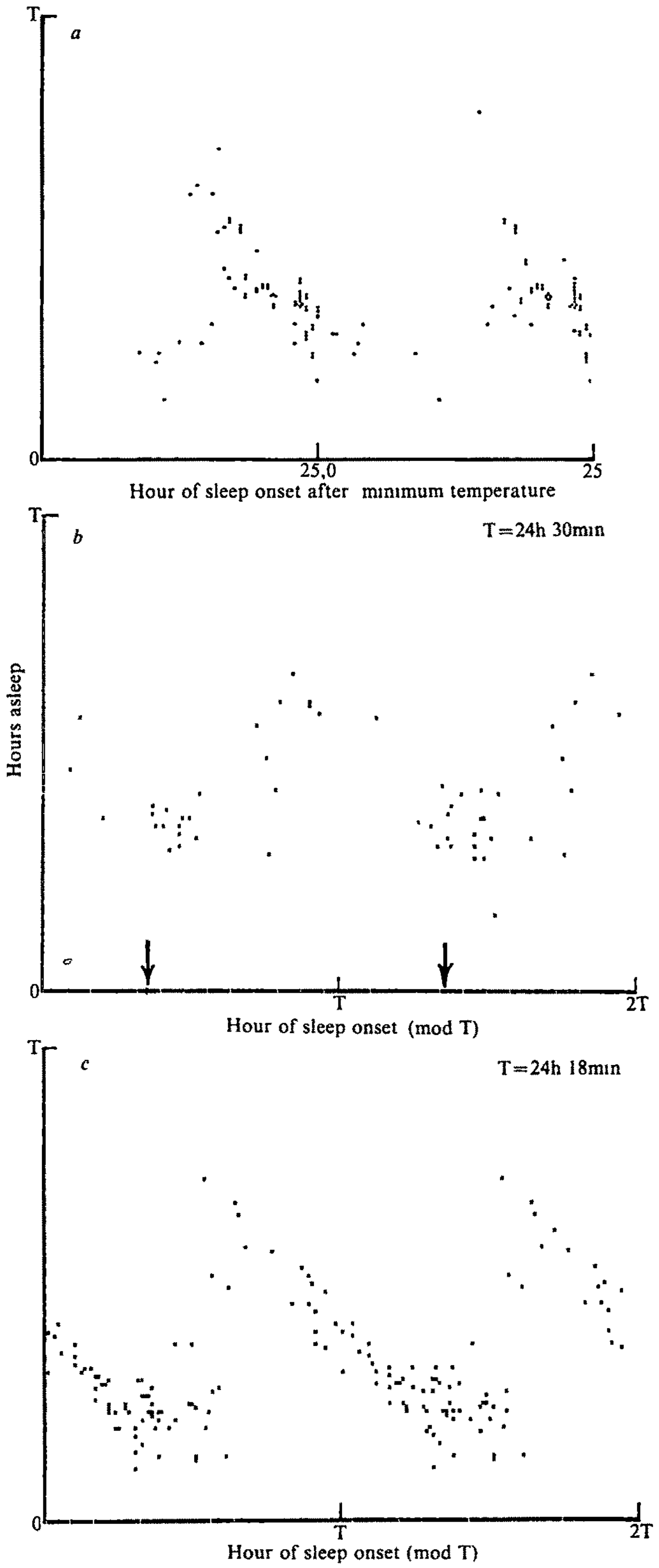


Fig. 2 *a*, Human sleep durations plotted against circadian phase (from ref 27) as described in the text *b*, As in *a*, with PRO1 data from ref 23, as described in text, the arrows mark approximate temperature minima *c*, As in *a* and *b* but data from refs 21, 22, subject J C digitized and reprocessed following Czeisler[23] The blank line inserted after 7 weeks represents transition to continuous dim light for the final 80 days Note that the vertical and horizontal scales are not identical here

within himself he just thinks that the room lights are being operated capriciously[11,36 89]

Some kinds of insomnia, previously dismissed as personal idiosyncrasy, seem to reflect treatable aberrations of the circadian clock and/or the sleep/wake system A simple manoeuvre is reported[90–92] to cure the 'delayed sleep phase syndrome', estimated to afflict hundreds of thousands of people Some people's sleep/wake rhythm, subject to rhythmic driving both by the internal temperature oscillator and by an external civil day, adopts an inconvenient phase relative to the civil day In the cases that come to clinical attention, both sleep and waking are tenaciously delayed about 6 h relative to what is usual for most people This makes it virtually impossible for them to hold any job that requires alertness in the morning However, by delaying three more hours each day for 6 days, one eventually arrives at the desired schedule and it remains stable long after termination of the phase-shifting therapy (Fig 4) However, the problem recurs after working late, travelling too far westward too abruptly, or otherwise losing yet 6 hours more Then another 18-h delay must be arranged[90 91] It would appear that such individuals can travel abroad only by remaining on a home-base schedule during a brief excursion, or, if they must synchronize to the civil day abroad, by circumnavigating and crossing the International Dateline westward

Daylight savings, jet-lag and shift work

A small phase shift of this sort has been imposed on hundreds of millions of people twice annually since the implementation of Benjamin Franklin's plan for daylight-saving time The possible consequences were never considered and are not fully known Monk and Aplin[93] recently found that the time of retiring adjusts immediately but that the time of awakening (or

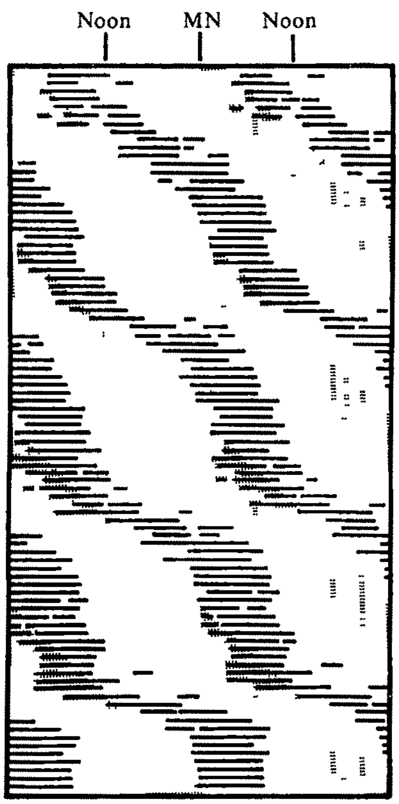


Fig. 3 In the same format as Fig 1 but with 24 0-h girth, a recording of more orderly alternations of sleep and waking in a non-blind individual who, nevertheless, does not synchronize to the 24-h world His own period waxes and wanes (the slope of the sleep-onset border varies) as his phase relation to the 24-h world changes (from ref 88) Note again the shaded zone of no-waking in this man's internal period, roughly corresponding to his temperature minimum

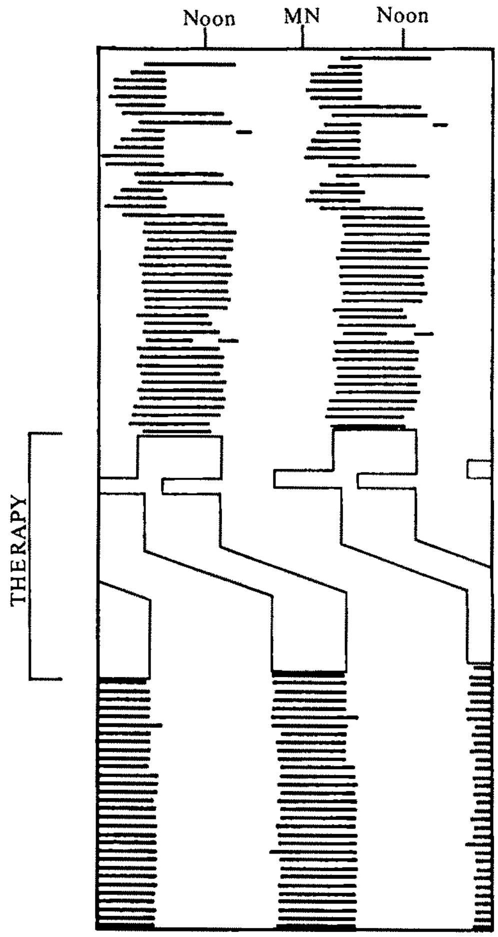


Fig. 4 The format is as in Fig 3 Sleep is indicated by a black bar in each of 110 stacked days During therapy, sleep was confined within the zigzag channel It retains the new timing with respect to the civil day (Adapted from ref 91)

duration of sleep, see Fig 2) needed a week to adjust fully The advancing (springtime) shift was somewhat more difficult than the delaying (autumnal) shift, as might be expected *a priori* for a species whose internal clock already has a somewhat longer period than the natural day Much the same happens when we travel either east or west Travel by air compresses the phase shift into a single day After transmeridian flights, about one person in four reports some indisposition associated with resynchronization[94] The abundant literature of jet-lag seems internally contradictory in many respects, and therefore suggests that individuals vary enormously in their reactions to diverse rescheduling stimuli Recent reviews[94–98] agree that internal phase relations remain abnormal for up to a week Nearly all flight experiments indicate substantially quicker adjustment after westward (delaying) flights across as many as nine time zones, than after comparable eastward (advancing) flights The latter also result in a depression of temperature-rhythm amplitude for several days Duration of flight, whether outbound or homebound, the time of departure and change in wake duration do not matter for the subsequent course of resynchronization

Both in laboratory isolation units and outside, the diverse physiological and psychological rhythms readjust at their diverse individual rates—even with some advancing and some delaying, most commonly when crossing 9 to 12 time zones[41 94] In free-running conditions of temporal isolation, subjects do not preceive the jet-lag-like dissociation of their own internal rhythms Wever[11 99] reports that subjects even feel better during the prolonged dissociation he calls internal desynchronization These observations conflict with universal expectation that

derangement of internal timing should cause discomfort or damage I know of no unequivocal demonstration that this internal dissociation in itself causes trouble apart from the inconvenience of trying to sleep, eat or think quickly at times when either local society or one's own body is not ready for it This latter discomfort is experienced by individuals who are out of synchrony with society but still internally synchronized This can be important shift workers include air-traffic controllers, nuclear submarine and power-plant operators, emergency-room doctors and intensive-care unit nurses Greater error rates have been measured during shift work[94,100,101] Individuals whose temperature exhibits larger daily excursions tend to take longer to adjust to each shift[41,99,102] This suggests itself as a selection criterion for individuals required to cope with frequent changes of scheduling

What is the best procedure for hastening adaptation to a new schedule (assuming this is a good thing to do)? Advice seems unreliable but the best guess may be just to maximize the contrast between our circumstances day and night with respect to light/dark, meals and social factors that provide synchronizing cues My personal impression, based on no more than conversation with other travellers during about fifty Atlantic crossings, is that for some people direct sunlight helps a lot

The timing of sleep and wakefulness is only one facet of the intricate patterning of temporal contrasts that pervade human physiology The medical profession is beginning to find ways to incorporate recent understanding of the predictable ups and downs of metabolism, the large excursions of hormone titre, drug susceptibility and plasma ion balance that previously limited the effectiveness and safety of routine procedures[103,104]

I am indebted for financial support during this survey to the NSF, the Institute for Natural Philosophy and the USAF School of Aerospace Medicine, and for hospitality to George Oster and the University of California at Berkeley during sabbatical leave from Purdue University I thank the many workers here cited who freely shared with me their time and unpublished data

1 Drucker-Colin, R *The Functions of Sleep* (eds Shkurovitch, M & Sterman, M B) (Academic, New York 1979)
2 Jones, H S & Oswald, I *Electroencephalography clin Neurophysiol* **24**, 378–380 (1968)
3 Meddis, R, Pearson, A J D & Langford, G *Electroencephalography clin Neurophysiol* **35**, 213–214 (1973)
4 Meddis, R *The Sleep Instinct* (Routledge & Kegan Paul, London, 1977)
5 Aschoff, J & Wever, R *Naturwissenschaften* **49**, 337–342 (1962)
6 Aschoff, J *Science* **148**, 1427–1432 (1965)
7 Siffre, M *Beyond Time* (Chatto & Windus, London, 1965)
8 Mills, J N *J Physiol, Lond* **174**, 217–231 (1964)
9 Findley, J D, Mialer, B M & Brady, J V in *Space Research Laboratory techn Rep Ser* (1963) (reproduced in ref 12)
10 Schaefer, K E, Clegg B R Carey, C R, Dougherty J H & Weybrew, B B *Aerospace Med* **38**, 1002–1018 (1967)
11 Wever R *The Circadian System of Man* (Springer, Heidelberg 1979)
12 Aschoff, J *Life Sci Space Res* **5**, 159–173 (North-Holland, Amsterdam, 1967)
13 Aschoff, J, Fatranska, M & Giedke, H *Science* **171**, 213–215 (1971)
14 Weitzman, E D, Czeisler, C A & Moore-Ede, M C in *Biological Rhythms and Their Central Mechanism* (eds Suda, M, Haishi, O & Nakagawa, H) (Elsevier, Amsterdam, 1979)
15 Wever, R *Int J Chronobiol* **3**, 19–55 (1975)
16 Weitzman, E D, Czeisler, C A, Zimmerman, J C & Ronda, J M in *Circadian and Ultradian Variations of Pituitary Hormones in Man* (Martinus Nijhoff, Belgium, in the press)
17 Weitzman, E D, Czeisler, C A, Zimmerman, J C & Moore-Ede, M C in *Neurosecretion and Brain Peptides* (eds Martin J B, Reichlin, S & Bick, K) 475–496 (Raven, New York, 1981)
18 Kronauer, R E, Czeisler, C A Pilato, S, Moore-Ede, M C & Weitzman, E D *Am J Physiol* **242**, R3–R17 (1982)
19 Kronauer, R E, Czeisler, C A, Pilato, S F, Moore-Ede, M C Weitzman, E D in *Proceedings Intra-science Symposium on New Perspectives in Sleep Research* (Santa Monica Symposium, in the press)
20 Kronauer, R E in *Mathematical Modeling of Circadian Systems* (eds Moore-Ede, M C & Czeisler, C A) (Raven, New York, 1982)
21 Jouvet, M, Mouret, J, Chouvet, G & Siffre, M in *The Neurosciences 3rd Study Program* (eds Schmitt, F & Worden, F) 491–497 (MIT Press, Cambridge, Massachusetts, 1974)
22 Chouvet, G, Mouret, J, Coindet, J, Siffre, M & Jouvet, M *Electroencephalography clin Neurophysiol* **37**, 367–380 (1974)
23 Czeisler, C A thesis, Stanford Univ (1978)
24 Zulley, J *Bericht über den* **30**, 398–399 (1976)
25 Zulley, J & Schulz, H in *4th Eur Cong Sleep Research*, Tirgu-Mures, 341–344 (Karger, Basel, 1980)
26 Zulley, J, Wever, R & Aschoff, J *Pflugers Arch ges Physiol* **391**, 314–318 (1981)
27 Zulley, J *Sleep* **2**, 344–346 (1980)
28 Zulley, J & Wever, R in *Vertebrate Circadian Systems Structure and Physiology* (eds Aschoff, J Daan S & Groos G) (Springer Berlin in the press)
29 Winfree, A T in *Mathematical Modeling of Circadian Systems* (eds Moore-Ede, M C & Czeisler C A) (Raven New York in the press)
30 Daan, S & Beersma, D in *Mathematical Modeling of Circadian Systems* (eds Moore-Ede, M C & Czeisler, C A) (Raven, New York, in the press)
31 Wever, R in *Mathematical Modeling of Circadian Systems* (eds Moore-Ede, M C & Czeisler, C A) (Raven, New York, in the press)
32 Borbely, A in *Curr Topics Neuroendocr* 1 (1981)
33 Eastman, C in *Mathematical Modeling of Circadian Systems* (eds Moore-Ede, M C & Czeisler, C A) (Raven New York, 1982)
34 Eastman, C in *Vertebrate Circadian Systems Structure and Function* (eds Groos, G A & Wehr, T A) (Springer, Berlin, 1982)
35 Buck, L *Aviat Space envir Med* **51**, 805–808 (1980)
36 Wever, R A *Rhythmical Aspects of Behavior* (in the press)
37 Czeisler C A, Richardson, G S, Zimmerman J C, Moore-Ede, M C & Weitzman, E D *Photochem Photobiol* **34**, 239–247 (1981)
38 Orth, D N & Island, D P *J clin Endocr Metab* **29**, 479–486 (1969)
39 Simenhoff, M L *J appl Physiol* **37**, 374–376 (1974)
40 Webb, W B, & Agnew, H W *Aerospace Med* **45**, 701–704 (1974)
41 Aschoff, J *Ergonomics* **21**, 739–754 (1978)
42 Nafarrate, A B & Miles, L E M *Proc XIV int Conf int Soc Chronobiol* (Il Ponte, Milan, 1982)
43 Minors, D S & Waterhouse, J M *Int J Chronobiol* **7**, 165–188 (1981)
44 Sulzman, F M, Fuller, C A & Moore-Ede, M C *J appl Physiol* **43**, 795–800 (1977)
45 Sulzman, F M, Fuller, C A & Moore-Ede, M, C *Physiol Behav* **18**, 775–779 (1977)
46 Sulzman, F M, Fuller, C A & Moore-Ede, M C *Am J Physiol* **234**, R130–R135 (1978)
47 Ehret, C F, Groh, K R & Meinert, H C *Adv exp Med Biol* **108**, 185–213 (1978)
48 Graeber, R C, Cuthbert, B N, Sing H L, Schneider, R J & Sessions, G R *Proc XIV int Conf int Soc Chronobiol* (Il Ponte, Milan, 1979)
49 Block, G D & Page, T L *A Rev Neurosci* **1**, 19–34 (1978)
50 Menaker, M, Takahashi, J S & Eskin, A *A Rev Physiol* **40**, 501–526 (1978)
51 Kawamura, H & Ibuka, N *Chronobiologica* **5**, 69–88 (1978)
52 Rusak, B & Zucker, I *Physiol Rev* **51**, 449–526 (1979)
53 Suda, M, Hayaishi, O & Nakagawa, H *Biological Rhythms and Their Central Mechanism* (Elsevier Amsterdam 1979)
54 Aschoff, J, Groos, G A & Wehr, T A *Vertebrate Circadian Systems Structure and Physiology* (Springer, Berlin, 1982)
55 Moore, R Y & Eichler, V B *Brain Res* **42**, 201–206 (1972)
56 Moore, R Y in *The Neurosciences Third Study Program* (eds Schmitt, F O & Worden, F G) 537–542 (MIT Press, Cambridge, Massachusetts, 1974)
57 Moore, R Y *Front Neuroendocr* **5**, 185–206 (1978)
58 Stephan, F K & Zucker, I *Physiol Behav* **8**, 315–326 (1972)
59 Stephan, F K & Zucker, I *Proc natn Acad Sci U S A* **69**, 1583–1586 (1972)
60 Lydic, R, Schoene, W C, Czeisler, C A & Moore-Ede, M C *Sleep* **2**, 355–361 (1980)
61 Nichino, H, Koizumi, K & Brooks, C M *Brain Res* **112**, 45–59 (1976)
62 Groos, G A & Mason, R *Neurosci Lett* **8**, 59–64 (1978)
63 Groos, G A & Mason R *J comp Physiol* **135**, 349–356 (1980)
64 Sawaki, Y *Expl Brain Res* **37**, 127–138 (1979)
65 Mai, J K *J Hirnforsch* **19**, 213–288 (1978)
66 Groos, G A Preprint (1981)
67 Inouye, S T & Kawamura, H *Proc natn Acad Sci U S A* **76**, 5962–5966 (1979)
68 Scwartz, W J & Gainer, H *Science* **197**, 1089–1091 (1977)
69 Schwartz, W J Smith, C B & Davidsen, L in *Biological Rhythms and Their Central Mechanism* (eds Suda M, Hayaishi, O & Nakagawa, H) 355–364 (Elsevier, Amsterdam, 1979)
70 Schwartz, W J, Davidson, L C & Smith, C B *J comp Neurol* **189**, 157–167 (1980)
71 Perlow, M J, Reppert, S M, Tamarkin, L Wyatt, R J & Klein D C *Brain Res* **182**, 211–216 (1980)
72 Richter, C P *Johns Hopkins med J* **141**, 47–61 (1977)
73 Lewy, A J, Wehr, T A, Goodwin, F K, Newsome, D A & Markey, S P *Science* **210**, 1267–1269 (1980)
74 Lewy, A J, Wehr, T A, Goodwin, F K, Newsome, D A & Rosenthal, N E *Lancet* **i**, 383–384 (1981)
75 Klein, K E & Wegmann, H M in *Chronobiology* (eds Scheving, L, Halberg, F & Pauly, J) 564–570 (Thieme, Stuttgart, 1974)
76 Halberg, F *Symp Bel-Air III*, 73–126 (Masson & Cie, Geneva, 1968)
77 Dirlich, G, Kammerlo, A, Schulz H, Lund, R, Doerr, P & Vonzerss, P *Biol Psychiat* **16**, 163–179 (1981)
78 Wehr, T A, Wirz-Justice, A, Goodwin, F K, Duncan, W & Gillin, J C *Science* **206**, 710–713 (1979)
79 Wehr, T A, Goodwin, F K, Wirz-Justice, A & Craig, C *Archs gen Psychiat* (in the press)
80 Wehr, T A & Goodwin, F K *Am Handbk Psychiat* **7**, 46–74 (1981)
81 Lewy, A J, Kern, H E, Rosenthal, N E & Wehr, T A, Goodwin, F K *Am J Psychiat* (submitted)
82 Lewy, A J & Newsome, D A *New Engl J Med* (submitted)
83 Pauly, J E Scheving, L, Burns E R Landon, J & Stone, J E *Proc XII int Conf int Soc Chronobiol* 19–28 (Il Ponte, Milan, 1977)
84 Sieber, W thesis, Univ Munich (1976)
85 Elliott, A L, Mills, J N & Waterhouse, J M *J Physiol, Lond* **212**, 30–31 (1971)
86 Weber, A L, Keyes, P, Conner, N & Cary, M S *Sleep* **2**, 347–354 (1980)
87 Miles L E M, Raymal, D M & Wilson, M R *Science* **198**, 421–423 (1977)
88 Kokkoris, C P *et al Sleep* **1**, 177–190 (1978)
89 Wever, R *Pfluger's Arch* (submitted)
90 Weitzman E D *et al Archs gen Psychiat* **38**, 737–746 (1981)
91 Czeisler, C A *et al Sleep* **4**, 1–21 (1981)
92 Moore-Ede M C *The Clocks That Time Us Structure and Function of the Circadian Timing System* (Harvard University Press 1982)
93 Monk, T H & Aplin, L C *Ergonomics* **23**, 167–178 (1980)
94 Klein, K E & Wegmann, H M in *NATO-AGARD-AG-247* (Technical Editing and Reproduction Ltd, London, 1980)
95 Aschoff, J, Hoffman, K, Pohl, H & Wever, R *Chronobiologica* **2**, 23–78 (1975)
96 Wever, R *Chronobiologica* **7**, 303–327 (1981)
97 Aschoff, J & Wever, R *Handbk behav Neurobiol* **4**, 311–331 (1981)
98 Hume, K I *Sleep* **2**, 417–435 (1980)
99 Wever, R in *Biological Rhythms in Sleep and Shift-work* (eds Johnson, L C, Tepas, D J, Colquhoun, W P & Colligan, M J) 35–60 (Spectrum, New York 1981)
100 Reinberg, A, Vieux, N & Andlauer, P *Adv Biosci* **30** (1981)
101 Johnson, L C, Tepas, D I, Colquhoun, W, C & Colligan, J, *Advances in Sleep Research* Vol 7 (Spectrum, New York, 1981)
102 Reinberg, A Vieux, N, Ghata J, Chaumont, A J & Laporte A *J Physiol, Paris* **74**, 405–409 (1978)
103 Scheving, L E & Pauly, J E *Chronobiologica* **1**, 3–21 (1974)
104 Halberg, F *Hosp Pract* **12**, 139–149 (1977)

ARTICLES

An explanation of $^{13}C/^{12}C$ variations in tree rings

R. J. Francey

CSIRO Division of Atmospheric Physics, Aspendale, Victoria, Australia 3195

G. D. Farquhar

Department of Environmental Biology, Research School of Biological Sciences, Australian National University, PO Box 475, Canberra City, ACT, Australia 2601

Variations in the $^{13}C/^{12}C$ ratio in trees are examined in the light of a simple expression relating the relative isotope composition of plant material, δ_p^{13} to δ_a^{13}, the atmospheric isotope value, c_a, the atmospheric CO_2 concentration and c_i, the internal concentration of CO_2 in leaves The expression gives good agreement with δ_p^{13} measurements where independent information on c_i exists, such as seasonal growth, growth low in the canopy and in conditions of low humidity. The expression provides possible explanations for two previously unexplained phenomena the absence of anticipated changes due to fossil fuel-induced changes in δ_a^{13}, and regional differences in δ_p^{13} trends

THE isotopic composition of carbon stored in the growth rings of trees may represent a record of $^{13}C/^{12}C$ variations in atmospheric CO_2, or of physiological responses to environmental changes, or a combination of both. In the first category, several authors[1–4] have attributed a decrease in $^{13}C/^{12}C$ in tree rings over the period 1850–1950 to an increasing component of ^{13}C-depleted CO_2, released into the atmosphere by increasing fossil fuel combustion, and/or by deforestation and expanding agriculture. In the second category, others[5–9] have sought an explanation of $^{13}C/^{12}C$ variations in tree rings through correlations with temperature Here there is normally an assumption of constant $^{13}C/^{12}C$ in the atmosphere, and a common, temperature-sensitive metabolic process within plants is implied In the third category, Mazany *et al*[10] assumed $^{13}C/^{12}C$ in air to be influenced by respiration from surrounding vegetation, which in turn responds to environmental factors such as temperature and water availability As in the first category Mazany *et al.* assumed constant fractionation in the carbon flow from the atmosphere to the biosphere

We investigate here the applicability to tree-ring studies of a new, quantitative expression[11], relating the carbon isotope ratio in those plants having the most common C_3 mechanism of photosynthetic carbon fixation to the physiological properties of a leaf When applied to tree-ring sequences corresponding to the period of industrialization, the fractionation expression carries implications concerning the response of trees to increasing levels of CO_2 in the atmosphere

Carbon discrimination in plants

An expression[11] for δ_p^{13} of the photosynthate in a C_3 plant is approximated in the form

$$\delta_p^{13} \simeq \delta_a^{13} - a - (b-a)c_i/c_a \qquad (1)$$

where δ_a^{13} is the fractional difference between $^{13}C/^{12}C$ in air surrounding the leaf and that in a reference material, c_a and c_i are the CO_2 concentrations in the external atmosphere and leaf intercellular spaces, respectively Furthermore,

$$c_i = c_a - A/g \qquad (2)$$

where A is the rate of CO_2 assimilation by the plant and g is the conductance of the boundary layer and stomatal pores to the diffusion of CO_2 The diffusivity of $^{13}CO_2$ in air is 4 4‰ less than that of $^{12}CO_2$, and this is the value of the constant a The value of b is determined by the isotopic discrimination of the ribulose-1, 5-bisphosphate carboxylase enzyme, thought to be the main source of fractionation in C_3 plants[12], with magnitude ~30‰ (ref 13)

In the derivation of equation (1) the CO_2 concentration drop from the intercellular spaces to the chloroplasts is assumed to be small, and intrinsic fractionation in photo- and dark-respiration steps is assumed negligible In its application to tree rings, the fractionation associated with translocation, and with the conversion of the photosynthate to other chemical factions such as cellulose or lignin, is assumed to be relatively small and not influenced by environmental factors In the application of equation (1) in this article we assume that b is constant

Qualitatively, this model is quite similar to those of Park and Epstein[12], Troughton[14] and Grinstead[15] Progress towards a quantitative model has also been made independently[13 16] but the ease of application of equation (1), is a significant advantage in interpreting carbon isotope behaviour in plants It is supported by the range of δ_p^{13} values observed in C_3 plants and by a correlation between δ_p^{13} and c_i/c_a in these species[17] Further support is drawn from measurements of δ_p^{13} and c_i/c_a following water stress[18] and of the temperature[19 20] and humidity[21] dependences of δ_p^{13} in growth chamber experiments

Paraphrasing[19], factors which reduce CO_2 assimilation rate through effects primarily on the mesophyll capacity for photosynthesis (for example, very low light intensities, and deficiencies of certain mineral nutrients) will increase c_i/c_a and reduce δ_p^{13}, factors which reduce CO_2 assimilation rates primarily through reduction of supply of CO_2 through the stomata (for example, large vapour pressure deficits) will decrease c_i and increase δ_p^{13} Note that in some situations a change in an external parameter may affect both A and g in equation (2) with no marked effect on c_i/c_a and thus δ_p^{13} (ref. 19)

Discussion

The possible relevance of equation (1) is explored with reference to observations of a number of reported phenomena Note that because relative rather than absolute δ^{13} values are discussed, the term δ_p^{13} is used at different times in reference to different material (such as cellulose, whole wood, and leaves) We discuss (1) the seasonal variation in δ_p^{13}, (2) the 'juvenile' effect in δ_p^{13}, (3) correlations between δ_p^{13} and regional ring width, temperature and precipitation, (4) variations of δ_p^{13} within a tree, (5) the absence of a fossil fuel-induced depletion of ^{13}C in Tasmanian trees, (6) regional differences in long-term δ_p^{13} trends

0028-0836/82/180028—04$01 00

Seasonal variations in δ_p^{13}

Lowden and Dyck measured a 2–5‰ decrease in δ_p^{13} values in both maple leaves and grass collected from an exposed site in Quebec during May to September The seasonal (and diurnal) variation in δ_a^{13} in the free atmosphere obeys a simple atmosphere biosphere mixing relationship[23–25] With typical ≥ 10 p p m v concentration decreases from May to September at high northern latitudes, this implies a δ_a^{13} increase of ~0 5‰, far smaller and in an opposite sense to the seasonal change in δ_p^{13} Lowden and Dyck hypothesize that the δ_p^{13} changes might be a direct consequence of changing photosynthesis to photorespiration ratios

Equation (1) is a quantitative description of such processes Whilst no information has been located on the seasonal variation of c_i in trees, there is ample evidence in measurements on crop plants of c_i increasing with age from peak photosynthetic rate to near senescence[26–30] Typical increases are 20–50 p p m v , which in equation (1) for typical values of c_a, δ_a and b, imply δ_p^{13} decreases of 2–4‰, in close agreement with the values of Lowden and Dyck in both leaves and grass

During initial leaf development A will be low and c_i high, that is over the full growth season we might expect a rapid decrease in c_i, a levelling off, followed by the slow increase to senescence[30] If this occurs in *Pinus radiata*, then measurements of 1–2‰ changes in δ_p^{13} within a ring of a New Zealand specimen[31] can be readily explained by equation (1) These indicate increasing δ_p^{13} early in the growth season (1/6th of a ring width), with high or slowly decreasing δ_p^{13} over the middle (4/6ths of a ring), and a very low δ_p^{13} in the late wood (1/6th of a ring) The effect is measured in both cellulose and lignin, confirming a mechanism early in the sequence of CO_2 assimilation The trend of the δ_p^{13} change with respect to season is confirmed by measurements on early and late wood from Australian *Athrotaxis selaginoides*[32] (At these latitudes c_a and δ_a^{13} in the free atmosphere are practically constant throughout the growth season[33 34]). In both cases the magnitude of the δ_p^{13} change of 1–2‰ is perhaps too large to be explained by temperature dependent fractionation process, or by local modification of δ_a^{13}, but quite compatible with observed seasonal changes in c_i

The juvenile effect in δ_p^{13}

Steadily increasing δ_p^{13} values with time, often observed in the innermost rings of a tree[4,15 35 36], have been attributed[4 37] to reassimilation of respired CO_2 which is retained within the plant canopy But this explanation has been questioned for trees which grew in relative isolation, and an age-related physiological factor suggested[15 35]

A step increase in δ_p^{13} accompanying a step increase in ring width[38] during early growth years, together with no δ_p^{13} increase in a well-exposed island tree[38], argue against an intrinsic age-related effect, and are qualitatively compatible with a 'canopy-effect'

The close proportionality between δ_a^{13} and c_a values observed in forest environments by Keeling[24 25] can be used to predict c_a levels some 40 p p m v above ambient, which must be maintained on average during photosynthesizing periods, to produce the typical 2‰ magnitude of the 'juvenile effect' Studies in a coniferous forest[39] (and O T Denmead, unpublished) and in a tropical rain forest[40] show that during daytime CO_2 in air is generally depleted by <4 p p m v through most of the stand, higher concentrations only occur close to the forest floor Even in dense forests δ_a^{13} is similar to that in background air[24 25] Periods of maximum photosynthesis generally coincide with periods of maximum instability and convective mixing in the atmospheric surface layer, whereas the value of 40 p p m v above ambient represents night-time conditions of extreme stability and dark-respiration in the Keeling data, and far exceeds any daytime value Normal selection of trees for δ_p^{13} analysis minimizes those situations of extreme atmospheric stagnation which might favour significant ^{12}C cyclic enrichment in trees[37]

While soil respiration may contribute to low δ_p^{13} low in the canopy through δ_a^{13} in equation (1), changes in c_i/c_a offer an alternative, or at least a contributory, explanation At low irradiances rate of assimilation, A, and to a lesser extent, stomatal conductance, g, are reduced The net result is that c_i increases[41]. In the extreme, as the irradiance drops near that level at which the photosynthetic rate is zero, c_i approaches c_a, the concentration of CO_2 outside the leaf On these grounds we expect low δ_p^{13} values at low irradiances.

Measurements on wheat plants grown in well-ventilated controlled chambers under shade cloth (in 220 $\mu E\ m^{-2}s^{-1}$) when compared with similar plants grown in light levels of 640 $\mu E\ m^{-2}s^{-1}$ show a 2‰ reduction of δ_p^{13} (ref 20), strongly suggesting that low light levels do produce δ_p^{13} changes of the sign and magnitude of the observed juvenile effect

Shading is unlikely to produce an explanation for juvenile ^{13}C depletion in isolated trees However, there are other external factors which affect the mesophyll capacity for photosynthesis and which could well vary during early tree development, for example, certain nutrient deficiencies, so that equation (1) provides a promising starting point for further research

δ_p^{13} and ring width, temperature and precipitation

The seasonal variation interpretation, plus interpretation of δ_p^{13} versus temperature results from growth chamber measurements[19,20], associate more positive δ_p^{13} with faster assimilation rate, A Correlations of δ_p^{13} with ring width by Mazany *et al*[10] and Tans and Mook[42] associate more positive δ_p^{13} with narrow rings and thus presumably slower assimilation rates

The general response of trees in arid areas is for ring widths to be wide in wet, cool years and to be narrow in warm, dry years, extreme water stress can cause stomatal closure and reduced CO_2 uptake as shown by Fritts[43] Low humidity has been shown[44] to reduce g and c_i, so that in conditions of high temperature and low precipitation we could expect narrower rings, lower c_i and, by application of equation (1), higher δ_p^{13}, as seen by Winter and colleagues[18 21] Measurements on juniper leaves and twigs from a wide area of arid Arizona, in which more negative δ_p^{13} are associated with lower growth temperatures and higher precipitation[45], are compatible with this

A similar argument can be applied to the Mazany *et al*[10] trees from the same area, however, they demonstrate a striking improvement in the δ_p^{13} versus ring-width index correlation when a regional rather than individual ring width index is used This seems to be their major reason for favouring an explanation in terms of regional modification of the atmosphere by respired CO_2 (that is, good conditions are reflected in vigorous regional growth, more respiration and an atmosphere with depleted δ_a^{13}) The rapidity with which CO_2 mixes in the atmosphere, particularly during photosynthesizing hours and in sparsely vegetated areas[24 25 46 47] would seem to invalidate an explanation involving significant regional modification of c_a, δ_a^{13} values In the context of equation (1), the higher correlation with regional indices requires that changes in average c_i are primarily determined by contemporary regional conditions whereas many factors are known to determine individual ring widths The fact that δ_p^{13} in the Dutch trees correlates inversely with ring width and directly with temperature and inversely with precipitation[42], as inferred for the Arizona trees, is suggestive that similar influences are at work

A recent summary of δ_p^{13} versus temperature coefficients indicates considerable complexity[48] In general, the δ_p^{13} response in tree rings to environmental parameters such as temperature and precipitation might be expected to be complex if δ_p^{13} is determined from equation (1) For example, growth chamber measurements on tree seedlings show how both positive and negative δ_p^{13} versus temperature coefficients might be obtained, depending on whether the change is below or above the temperature corresponding to peak photosynthetic

activity[19,20], and the optimum temperature is different for different species Thus there may be both a species and a range dependence Added to this will be quite non-linear influences resulting from stress, the most obvious being stomatal response to extreme hot dry conditions This places great emphasis on site and species selection if δ_p^{13} in trees is to be used for environmental reconstructions

Variations in δ_p^{13} within a tree

Several authors have drawn attention to variations of δ_p^{13} around the circumference of a ring[4,10,38,42] of magnitude 1–4‰ Again it is difficult to explain the magnitude of this effect in terms of local modification of δ_a^{13}—previous measurements of δ_a^{13} versus c_a (refs 24, 25) require large CO_2 gradients of 20–80 p p m v to be maintained across a tree for the major portion of the growth periods It is especially difficult as these trees have generally been selected from positions of relative isolation

For the same reason large differences due to a cyclic enrichment process[37] are difficult to envisage The main evidence presented by Keeling for the cyclic enrichment process, effectively a comparison of maximum daytime δ_a^{13} values with maximum night-time c_a values at various forest sites, can be explained by the seasonal variation in the large-scale atmospheric δ_a^{13} values

Variations in stomatal conductance, g, for example due to differences in structure and shading in leaves[49], or variations in leaf assimilation rate, A, due to shading[20], appear to offer a more realistic explanation through equations (1) and (2) For example, even though Wong *et al*[50] show that c_i is approximately constant in *Eucalyptus pauciflora* over a range of irradiances above 250 $\mu E\, m^{-2} s^{-1}$, the small changes observed are sufficient to imply δ_p^{13} changes of ~1‰

No fossil-fuel depletion of ^{13}C in Tasmanian trees

Cellulose from eight Tasmanian trees[38] exhibits a practically constant δ_p^{13} with time over the period 1750–1970 This is in contrast to the predicted decrease of 0 7–0 8‰ (refs 42, 51) There is even greater contrast if an additional large amount of CO_2 has been released as a result of forest clearing and agricultural practices[31] It does not reflect direct observation of a δ_a^{13} change in the atmosphere[52] It also contrasts with a summary of most reported Northern Hemisphere δ_p^{13} measurements[53]

The eight Tasmanian trees represent two species, and a wide range of environments (small island to extremely exposed mountain ridge) but with all sites sharing desirable features in the context of monitoring global atmospheric CO_2 The areas are remote from large-scale industry and experience persistent winds off the southern ocean, enhancing mixing The seasonal variation in CO_2 is small and temperatures are stabilized by the surrounding (particularly upwind) oceans The very high and persistent rainfall (typically in excess of 2,000 mm yr^{-1}) implies minimum water stress with its corresponding effects on fractionation In a forest situation the selected species typically have most of their foliage at or above canopy level

The absence of a δ_p^{13} trend which reflects the anticipated δ_a^{13} trend in any one of the eight trees, requires an effect operating on a regional scale (≥100 km) of similar magnitude but opposite sign to the fossil fuel signature in the atmosphere Equation (1) provides a possible solution without recourse to complete coincidence While the δ_a^{13} decrease due to fossil fuel release is associated with an increase in c_a, the effect of increasing c_a in the equation, all else being constant, is to increase δ_p^{13}

Physiological data on c_i/c_a in trees is scarce, and over the 50–100 year time scale, is non-existent The strength of the present argument rests with the feasibility of the mechanism, when no alternative explanations exist Even those physiological experiments which suggest constancy of c_i/c_a in some plants over short time scales[19], are of an accuracy which do not exclude the small changes necessary to reconcile the Tasmanian δ_p^{13} measurements with the anticipated δ_a^{13} changes due to fossil fuel combustion

For example, combination of the predicted δ_a^{13} curve[42,51] with the Tasmanian mean curve plus equation (1) implies

$$[c_i/c_a]_{1970} \simeq 0\ 9[c_i/c_a]_{1870}$$

and thus

$$[c_i]_{1970} \simeq [c_i]_{1870}$$

In other words the trees have, to a first approximation, maintained a constant c_i This conclusion does not depend on the value of b, and depends only weakly on the δ_p^{13} value used The simplistic application of equations (1) and (2) implies either reduced stomatal conductance, g, or increased assimilation rate A over the 100 years of increasing CO_2 levels In view of the abundant water availability, the latter is favoured

There are many complexities in the real situation which must be better understood before the model for fractionation and the tree records can be used with confidence in this fashion However, we have illustrated a possible sensitive means of probing physiological behaviour over long time scales

Regional differences in long-term δ_p^{13} trends

A composite δ_p^{13} curve from 26 European and North American trees shows a decrease of almost 2‰ from 1850 to 1975, which includes a slight increase between 1940 and 1960 (refs 4, 53) Several Northern Hemisphere individual trees[2,3] have been reported with similar δ_p^{13} decreases up to around 1940–50, all in marked contrast to the practically constant δ_p^{13} in the Tasmanian trees from 1850 to 1970

Considerable caution is required in attributing these differences in δ_p^{13} to environmental differences on a hemispheric scale For example, the mean of four Dutch trees[42] decreases between 1860 and 1940 by no more than 0 5‰, with an increase of almost 0 4‰ from 1940 to 1970, overall these are not significantly different (in view of the scatter between individual trees) from the Tasmanian trees But the one other reported Southern Hemisphere tree-ring result[1] is very different from the Tasmanian results with a steady 1–1 5‰ decrease from 1890 to 1960

The predicted change in global δ_a^{13} due to fossil fuel combustion of around 0 7–0 8‰ (refs 42, 51) from 1750 to 1970 is equally incompatible with the 2‰ change of Freyer[4] as it is with the flat Tasmanian record In both cases the differences from the predicted curve are too large to be explained by measurement error A large additional release of CO_2 around 1860 associated with an 'agricultural explosion' has been postulated to explain the larger δ_p^{13} trends[4,31]

Use of an 0 032‰ p p m v^{-1} biosphere–atmosphere mixing relationship observed[54] between 1956 and 1978 as an upper limit to that applicable over the past 200 yr can be used to predict 1750 levels of atmospheric CO_2 considerably lower than the 270–290 p p m v typically employed in global carbon budgeting models (G I Pearman and I Enting, personal communication) More realistic mixing over the 200-yr time scale comes from such models[42,52], which include realistic air–sea fractionation factors, and these suggest a relative insensitivity of δ_a^{13} to biospheric release Thus it is not obvious that the postulated explanation for the large δ_p^{13} trends is valid

The maximum latitudinal difference in δ_a^{13} that can be maintained over decades is expected to be small For example, the current difference in annual average CO_2 concentration of 3–4 p p m v from pole to pole[31], if assumed due to fossil fuel combustion, represents a 0 1–0 2‰ difference in δ_a^{13} Keeling *et al*[54] measured a difference between North America and South Pole of <0 1‰ in 1978, although results of Kroopnick *et al*[55] in 1970, when corrected for seasonal effects indicate a 0 5‰ difference over a similar latitude range A recent two-dimensional and seasonal model of global carbon exchange

(refs 33, 56–58 and P. Hyson, personal communication), estimate a total change, 1850–1970, of ~0.2‰ between 50°N and 50°S. Thus even if the Freyer[4] results do reflect a global δ_a^{13} change, the conflict with the Tasmanian δ_p^{13} trends is about twice that discussed above. It is true that both cannot represent large scale δ_a^{13} changes.

It is equally unrealistic to suggest that regional or sub-canopy modification of δ_a^{13} could explain the differences. This is because of the small magnitude of daytime δ_a^{13} differences from background air in even quite dense forests[24,25], and the wide variation in ages, exposures, and species, of trees showing common trends[38]. The only possibility of reconciling all observations, seem to be in systematic variations in the fractionation of atmospheric CO_2 as it is assimilated by trees.

Drawing on examples and discussion above, possible examples of environmental influences on the fractionation process in trees which operate locally, but on decadal to century time scales, include: (1) shading in developing forests, where individuals have significant subcanopy foliage; (2) nutrient depletion; (3) changes in assimilation rate resulting from regional temperature trends, or from interactions with changing atmospheric pollutant levels; (4) changes in stomatal behaviour resulting from trends in atmospheric pollutants or humidity.

For tree-ring sequences over the past 50–100 yr it can be argued that δ_a^{13} is known to greater accuracy than c_i/c_a; using the fractionation expression, measurements of δ_p^{13} may become useful in plant ecophysiology.

We thank Dr G. I. Pearman, Mr P. Hyson and Dr Ian Enting (CSIRO, Division of Atmospheric Physics) for computational work with respect to the carbon budgeting models. Professor C. B. Osmond and Dr Ian Cowan and Dr Roger Gifford offered valuable comments.

Received 7 January; accepted 1 March 1982.

1. Rebello, A. & Wagener, K. *Proc. int. Symp. on Environmental Biogeochemistry* Vol. 1, (ed. Nriagu, J. O.) 13–23 (Ann Arbor, Michigan, 1976).
2. Stuiver, M. *Science* **199**, 253–258 (1978).
3. Wilson, A. *Nature* **273**, 40–41 (1978).
4. Freyer, H. D. *Tellus* **31**, 124–137 (1979).
5. Pearman, G. I., Francey, R. J. & Fraser, P. J. B. *Nature* **260**, 771–773 (1976).
6. Farmer, J. G. *Nature* **279**, 229–231 (1979).
7. Lerman, J. C. *Int. CNRS* **219**, 163–191 (1974).
8. Libby, L. M. *et al. Nature* **261**, 284–288 (1976).
9. Grinstead, M. J., Wilson, A. T. & Ferguson, C. W. *Earth planet. Sci. Lett.* **42**, 251–253 (1979).
10. Mazany, T., Lerman, J. C. & Long, A. *Nature* **287**, 432–435 (1980).
11. Farquhar, G. D., O'Leary, M. H. & Berry, J. A. *Aust. J. Pl. Physiol.* **9**, 121–137 (1982).
12. Park, R. & Epstein, S. *Geochim. cosmochim. Acta* **21**, 110–123 (1960).
13. O'Leary, M. H. *Phytochemistry* **20**, 553–567 (1981).
14. Troughton, J. H. *Proc. 8th int. Radiocarbon Dating Conf.*, Lower Hutt, E40–E57 (1972).
15. Grinstead, M. J. thesis, Univ. Waikato (1977).
16. Vogel, J. C. *Sitzungsberichte der Heidelberger Akademie der Wissenschaften Mathematisch-naturwissenschaftliche Klase Jahrgang* (Springer, Berlin, 1980).
17. Farquhar, G. D. Ball, M. C. von Caemmerer, S. & Roksandic, S. *Oecologia* **52**, 121–124 (1982).
18. Winter, K. *Z. Pflanzenphysiol.* **101**, 421–430 (1981).
19. Farquhar, G. D. in *Carbon Dioxide and Climate: Australian Research* (ed. Pearman, G. I.) 105–110 (Australian Academy of Science, Canberra, 1980).
20. Francey, R. J. in *Carbon Dioxide and Climate: Australian Research* (ed. Pearman, G. I.) 95–104 (Australian Academy of Science, Canberra, 1980).
21. Winter, K., Holtum, J. A. M., Edwards, G. E. & O'Leary, M. H. *J. Exp. Bot.* (in the press).
22. Lowden, J. A. & Dyck, W. *Can. J. Earth Sci.* **11**, 79–88 (1974).
23. Keeling, C. D. *Tellus* **12**, 200–203 (1960).
24. Keeling, C. D. *Geochim cosmochim. Acta* **13**, 322–334 (1958).
25. Keeling, C. D. *Geochim. cosmochim. Acta* **24**, 277–298 (1961).
26. von Caemmerer, S. thesis, Australian National Univ. (1981).
27. Davis, S. D. & McCree, K. J. *Crop Sci.* **18**, 280–282 (1978).
28. Constable, G. A. & Rawson, H. M. *Aust. J. Pl. Physiol.* **7**, 89–100 (1980).
29. Rawson, H. M. & Constable, G. A. *Aust. J. Pl. Physiol.* **7**, 555–573 (1981).
30. Woodward, R. G. & Rawson, H. M. *Aust. J. Pl. Physiol.* **3**, 257–267, (1976).
31. Wilson, A. T. & Grinstead, M. J. *Nature* **265**, 133–135 (1977).
32. Fraser, P. J. B., Francey, R. J. & Pearman, G. I. *DSIR Bull.* **220**, 67–73 (1978).
33. Fraser, P. J., Hyson, P. & Pearman, G. I. in *Carbon Dioxide and Climate: Australian Research* (ed. Pearman, G. I.) 49–60 (Australian Academy of Science, Canberra, 1980).
34. Goodman, H. S. in *Carbon Dioxide and Climate: Australian Research* (ed. Pearman, G. I.) 111–114 (Australian Academy of Science, Canberra, 1980).
35. Craig, H. *Science* **119**, 141–143 (1954).
36. Jansen, H. S. *Nature* **196**, 84–85 (1962).
37. Keeling, C. D. *Geochim. cosmochim. Acta* **24**, 299–313 (1961).
38. Francey, R. J. *Nature* **290**, 232–235 (1981).
39. Jarvis, P. G., James, G. B. & Landsberg, J. J. in *Vegetation and the Atmosphere* Vol. 2 (ed. Monteith, J. L.) 171–240 (Academic, London, 1976).
40. Allen, L. H. Jr & Lemon, E. R. in *Vegetation and the Atmosphere* Vol. 2 (ed. Monteith, J. L.) 265–308 (Academic, London, 1976).
41. Smith, B. N., Oliver, J. & McMillan, C. *Bot. Gaz.* **137**, 99–104 (1976).
42. Tans, P. P. & Mook, W. G. *Tellus* **32**, 268–283 (1980).
43. Fritts, H. C. *Tree Rings and Climate*, 173–174 (Academic, London, 1976).
44. Farquhar, G. D., Schulze, E.-D. & Küppers, M. *Aust. J. Pl. Physiol.* **7**, 315–327 (1980).
45. Arnold, L. D. in *Short Papers of 4th int. Conf. on Geochronology and Isotope Geology* (ed. Zartman, R. E.) (US Geological Society, 1978).
46. Garratt, J. R. & Pearman, G. I. *Atmospheric Environment* Vol. 7, 1257–1266 (Pergamon, Oxford, 1973).
47. Pearman, G. I. & Garratt, J. R. *Agricul. Met.* **12**, 13–25 (1973).
48. Long, A. in *Climate from Tree Rings* (eds Hughes, M. K., Kelly, P. M., Pilcher, J. R. & La Marche, V. C. Jr) (Cambridge University Press, in the press).
49. Fritts, H. C. *Tree Rings and Climate*, 164 (Academic, London, 1976).
50. Wong, S. C., Cowan, I. R. & Farquhar, G. D. *Nature* **282**, 424–426 (1979).
51. Pearman, G. I. in *Carbon Dioxide and Climate: Australian Research* (ed. Pearman, G. I.) 11–20 (Australian Academy of Science, Canberra, 1980).
52. Keeling, C. D., Bacastow, R. B. & Tans, P. P. *Geophys. Res. Lett.* **7**, 505–508 (1980).
53. Freyer, H. D. *WMO Project on Research and Monitoring of Atmospheric CO_2* Rep. No. 1, Appendix E (1979).
54. Keeling, C. D., Mook, W. G. & Tans, P. P. *Nature* **277**, 121–123 (1979).
55. Kroopnick *et al.* (1977). Kroopnick P. M., Margolis, S. V. & Wong, C. S. in *The Fate of Fossil Fuel CO_2 in the Oceans* (eds Andersen, N. R. & Malahoff, A.) (Plenum, New York, 1977).
56. Pearman, G. I. & Hyson, P. *J. geophys. Res.* **85**, 4468–4474 (1980).
57. Hyson, P. & Pearman, G. I. in *Carbon Dioxide and Climate: Australian Research* (ed. Pearman, G. I.) 65–78 (Australian Academy of Science, Canberra, 1980).
58. Hyson, P., Fraser, P. J. & Pearman, G. I. *J. geophys. Res.* **85**, 4443–4455 (1980).

Function of a tRNA gene promoter depends on nucleosome position

Stephanie Wittig & Burghardt Wittig

Institut für Molekularbiologie und Biochemie, Freie Universität Berlin, Arnimallee 22, 1000 Berlin 33, FRG

We have used an in vitro *procedure to direct nucleosomes to different sites on a chicken embryo tRNA gene. The results show that transcription of the tRNA gene depends on nucleosome position and in a functional promoter structure, the constant gene sequence TTCGA coincides with the nucleosome middle axis.*

EVIDENCE is accumulating that the chromatin structure of expressed genes differs in several respects from non-expressed genes[1–14]. As genes are associated with nucleosomes regardless of their functional state[14], the recognition of nucleosomes that have become specifically altered requires a higher level of programming such as exposure of a recognition DNA base sequence. The particular location of such a sequence within the nucleosome could serve as the structure that is recognized. Such a nonrandom association of nucleosomes with respect to DNA base sequence is generally called 'phasing' although this term covers several different phenomena (for review see ref. 15). The limits of long-range phasing of regularly spaced nucleosomes and of sequence-specific nucleosome positioning have been discussed critically by Kornberg[16]. It is possible that base-specific DNA binding proteins serve as a phase trigger for regularly spaced nucleosomes. In the assays normally performed, such nucleosomes would display a nonrandom association with the DNA base sequence, unless due to inherent

0028–0836/82/180031—08$01.00

inaccuracies of nucleosome spacing, the effect was not apparent within a short distance from the phase trigger[16]. Recently, we have found such a short-range phasing of nucleosomes for several chicken embryonic tRNA genes[17]. The tRNA structural gene sequence always begins ~20 base pairs (bp) inside the nucleosome core, and the phased association probably does not extend farther than about five nucleosomes. In one fragment of nucleosomal DNA investigated, in which a second tRNA gene is present in the same orientation as, but does not fit into the nucleosome phasing of the first, the phase seems to be re-adjusted for the second gene. A strikingly similar picture was obtained by investigators who performed a computer simulation of the above—they found that a sequence-specific DNA binding protein triggers the phase for regularly spaced nucleosomes positioned essentially at random[16].

Our own view that the histones of the nucleosome cannot account for the phase relationship between tRNA genes and nucleosomes was supported in the early stages of the investigations described here. Phasing could not be restored in nucleosome assembly experiments using cloned chicken embryo $tRNA_2^{Lys}$ genes regardless of whether isolated histones[18], histone octamers[19] or nucleosome monomers served as the histone source[20]. Even when the cloned DNA was incubated before the addition of histones plus crude protein extracts to allow binding of sequence-specific proteins, the assembled nucleosomes were not phased. These results led us to hypothesize that the observed phase relationship between tRNA genes and nucleosomes may be established during replication, when DNA is newly packed into the repeating nucleosome structure. Sequence-specific DNA binding proteins involved in the replication process[21] and/or stretches of single-stranded DNA (ssDNA)[22,23] may be the phase-triggering structures. The *in vitro* system used here, which is described only briefly (details will be published elsewhere[24]) allows a site-directed assembly of nucleosomes. Using this system we investigated whether a particular location of a regulatory gene sequence within the nucleosome structure was suitable for programming the expression of that gene.

The DNA base sequences essential for promotion of tRNA gene transcription have been recently identified[25]. As in the case of genes encoding 5S RNA[26,27] these sequences are also located within the structural gene. They are split into two regions, one consisting of nucleotides constantly found in the dihydrouridine stem of mature tRNA, the other encoding the conserved loop sequence TTCGA of the tRNA pseudouridine stem. The latter sequence, which is the region of symmetry for a complementary palindrome[28], is localized *in vivo* near the middle axis of the nucleosome in chicken embryo tRNA genes.

The site-directed nucleosome assembly system

The assembly extract used was a crude nuclear preparation of chicken embryos[29], consisting of two components, a soluble nucleoprotein fraction and a histone source. The former fraction was prepared by incubating nuclei at elevated temperatures in an alkaline buffer, followed by sonication at high salt concentration. The procedure was identical to that used for the initial

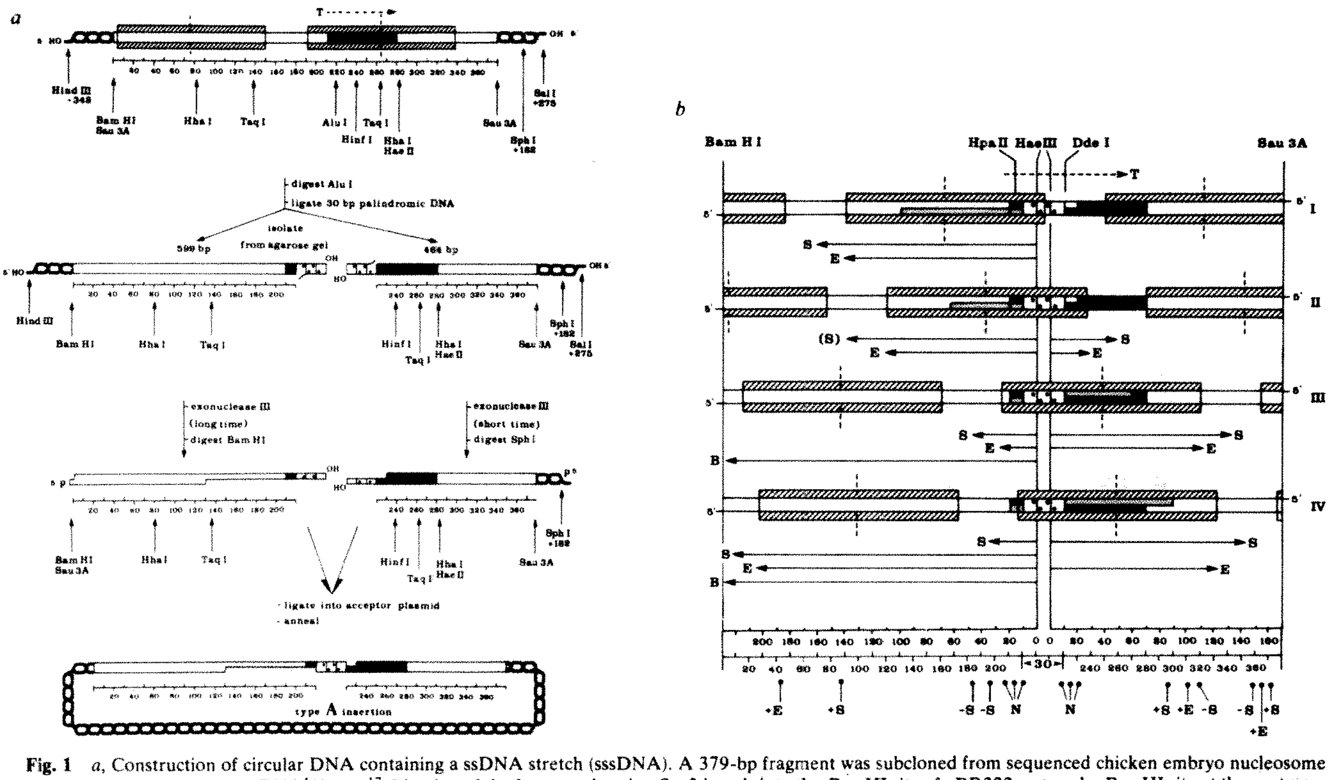


Fig. 1 *a*, Construction of circular DNA containing a ssDNA stretch (sssDNA). A 379-bp fragment was subcloned from sequenced chicken embryo nucleosome tetramer DNA containing a $tRNA_2^{Lys}$ gene[17]. Ligation of the fragment bearing *Sau*3A ends into the *Bam*HI site of pBR322 restored a *Bam*HI site at the upstream boundary of the fragment and left a *Sau*3A site at the downstream boundary. In the subclone used (pBR322 N4 *Bam Sau*) the tRNA gene is oriented clockwise with respect to the tRNA identical strand. Relevant restriction sites within the fragment are shown on a linear scale corresponding to the numbering of the DNA base sequence; restriction sites belonging to the plasmid vector are off the scale. Negative and positive nucleotide numbers indicate the respective distance from the original *Bam*HI site of pBR322. Open box, chicken embryo nucleosomal DNA; solid box, tRNA structural gene; dotted box, 30-bp palindromic DNA (trimer of synthetic *Hind*III linkers). Dots indicate position of ^{32}P-label; cross-hatched boxes, nucleosome cores; helical band, plasmid DNA. *b*, Position of assembled nucleosomes with respect to the tRNA gene. Assembled nucleosomes were either partially digested with *Staphylococcus* nuclease alone, or with *Staphylococcus* nuclease followed by exonuclease III (single strands were trimmed with S_1 nuclease in the isolated DNA). Nucleosome tetramer DNA was isolated by agarose gel electrophoresis, and eluted from the gel[43]. One half of the DNA was restricted with *Hae*III, *Dde*I and *Bam*HI, for the analysis of upstream fragments, then the other half was restricted with *Hae*III, *Hpa*II and *Sau*3A to obtain the downstream fragments. Fragments > 20 bp are single end-labelled radioactively (2.5×10^5 c.p.m. μg^{-1}). Fragments were analysed on long sequencing gels. Assembled nucleosomes were also digested with *Bam*HI, the restricted DNA isolated by Sepharose 2B chromatography and restricted with *Hae*III and *Dde*I. Fragments were analysed by agarose gel electrophoresis. I to IV, four sssDNA samples used for nucleosome assembly. Arrows labelled S, E and B represent fragments generated by *Staphylococcus* nuclease, exonuclease III and *Bam*HI, respectively; *T* indicates the direction of transcription. Dots labelled S represent sequences cleaved preferentially (+) or with low frequency (−) in isolated DNA by *Staphylococcus* nuclease; +E indicates pauses of exonuclease III on isolated DNA; N indicates the location of single-stranded nicks. Shaded area, original location of the ssDNA stretch in sssDNA repaired to double-stranded DNA during nucleosome assembly. Vertical broken lines, nucleosome middle axis.

a

5' GATCCCCTTTTGGGGTTTCATTTTGGGGTTTTTTTTGGGGGGGGGGGGGGGGTTCCTGCAATGCACCTTCCTCCCGCTGCGGTGTCTATGGTATAATAG 100

GAGCGACTGTATGAAGCAAATTCTGGCACTCATAAAATCGAGCCTTCCGTAAACCAGCCCCAAATCTGTAACTTAAGGCAGAAAAAAAGGCAGAAAAAAG 200

ACTCCAAGCTTGGCCAAGCTTGGCCAAGCTTGG (B)

CCAAGCTTGGCCAAGCTTGGCCAAGCTTGG (A)

GCTTTTTTTGCCCGGCTAGC TCAGTTGGTAGAGCATGAGACTCTTAATCTCAGGGTCGTGGGTTCGAGCCCCACGTTGGGCGCCCTTGGTTTTCGCCATC 300

TCACTTTTTGCCCCCAAAAAGGCACATTGCAAATGGCTGCTTGTCTCTATACTATCCCTTTTTCCCCCGCTATAGGGCT 3'

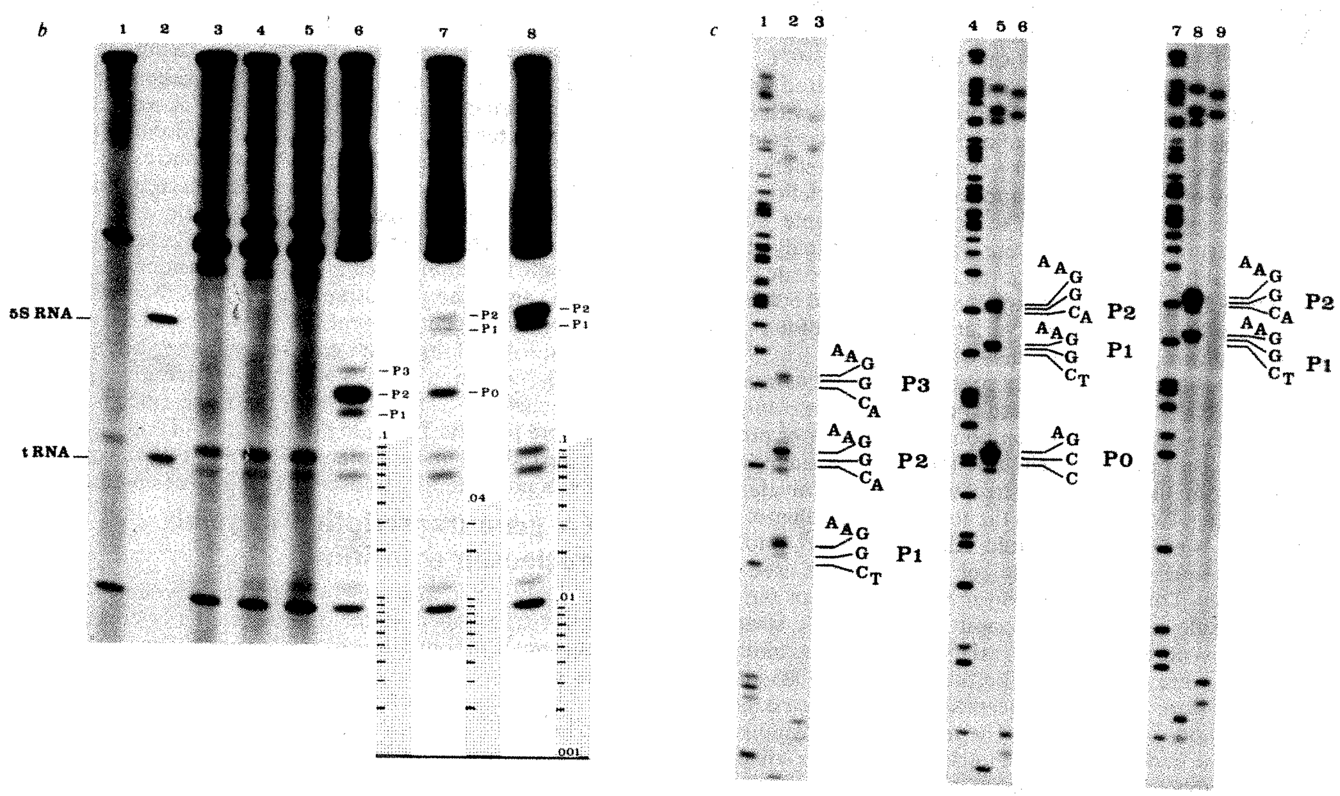


Fig. 2 *a*, DNA base sequence of the 379-bp subclone of a chicken embryo tRNA$_2^{Lys}$ gene. Broken line represents tRNA structural gene sequence; solid lines, 30-bp palindromic DNA. Vertical arrows labelled A and B indicate type A and type B insertions, respectively. *b*, Transcription of the tRNA$_2^{Lys}$ gene within randomly positioned nucleosomes. 1, Assembly extract minus sssDNA; 2, chicken embryo tRNA$_2^{Lys}$ and *E. coli* 5S RNA; 3, isolated DNA lacking 30-bp palindromic insertion (after incubation with assembly extract); 4, isolated DNA with type A insertion; 5, isolated DNA with type B insertion; 6, randomly positioned nucleosomes on DNA without insertion; 7, randomly positioned nucleosomes on DNA with type A insertion; 8, randomly positioned nucleosomes on DNA with type B insertion. The bands representing radioactively labelled tRNA precursors were cut out and quantified by liquid scintillation counting. Efficiencies of transcription (mol tRNA precursor per mol tRNA gene) are indicated by columns on a logarithmic scale shown beside the respective electrophoresis slot. A 25-μl nucleosome assembly extract was diluted to 500 μl with 100 mM KCl, 20 mM NgCl, 10 mM 2-mercaptoethanol, 10% glycerol, 50 mM Tris-HCl *p*H 7.9, and centrifuged at 4 °C in a type 25 rotor (Beckman Instruments) at 90,000*g* for 3 h. The supernatant was pipetted off and immediately 100 μl of 100 mM KCl, 20 mM NaCl, 2 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 10% glycerol, 0.5 mM ATP, 0.5 mM CTP, 0.5 mM UTP, 50 mM Tris-HCl *p*H 7.9, were added and vortexed briefly. Of 0.25 μg sssDNA used in the assembly extract, at least 70% were recovered (as monitored by the ^{32}P-label in the 30-bp palindromic DNA) together with ~50 μg (5%) and ~2 μg (3%) of residual protein and nucleic acids, respectively, from the assembly extract. 12.5 μl of 250 μM [α-^{32}P]GTP (10 Ci mmol^{-1}) and 12.5 μl of partially purified RNA polymerase III (0.1 U, 20 μg protein) were added and incubated at 37 °C for 100 min. 'Carrier' plasmid DNA (5 μg) was added, and nucleic acids were deproteinized and precipitated, then dissolved in 25 μl of 1 mM EDTA, 10 mM Tris *p*H 7.4, and restricted with *Hin*dIII to cut ^{32}P-labelled DNA (originating from sssDNA) into 10-bp fragments. ^{32}P-RNA was analysed on 6% polyacrylamide, 7 M urea gels (2×250×450 mm). Chicken embryo DNA-dependent RNA polymerase III was purified as described elsewhere[30], and the material obtained after DEAE-Sephadex chromatography was used. *c*, S_1 nuclease hybridization mapping of transcription products. 1, Guanosine-specific partial chemical cleavage (G-ladder) of sense transcription probe from DNA without insertion; 2, randomly positioned nucleosomes on DNA without insertion, sense transcription mapping; 3, as in 2, but anti-sense transcription mapping; 4, G-ladder of sense transcription probe from DNA with type A insertion; 5, experiment IV, sense transcription mapping; 6, expt IV, anti-sense transcription mapping; 7, G-ladder of sense transcription probe from DNA with type B insertion; 8, expt VIII, sense transcription mapping; 9, expt VIII, anti-sense transcription mapping. 1 μg of sssDNA was used in the assembly extract and transcription was performed in the presence of 0.5 mM GTP instead of ^{32}P-labelled GTP. After *Hin*dIII restriction, nucleic acids were precipitated and dissolved in 25 μl of 80% formamide, 0.4 M NaCl, 1 mM EDTA, 50 mM PIPES, *p*H 6.8 containing 1 μg of ^{32}P-labelled DNA hybridization probe. Hybridization was at 55 °C for 16 h. 250 μl of ice-cold 200 mM NaCl, 1 mM $ZnSO_4$, 25 mM Na-acetate *p*H 4.5 and 10 U of S_1 nuclease were added and incubated at 37 °C for 3 h. Nucleic acids were precipitated (5 μg of carrier tRNA were added) and analysed on a 10% sequencing gel. Probes for S_1 nuclease hybridization mapping were prepared from plasmids pBR322 N4 *Bam SauA* and pBR322 N4 *Bam SauB*, containing the 379-bp fragment of nucleosomal DNA (see Fig. 2*a*) with a type A and type B insertion, respectively.

solubilization of class I, II and III DNA-dependent RNA polymerases from chicken embryos[30]. A nucleosome monomer fraction comprising a mixture of core particles and nucleosomes (the histone H1 content being reduced to <20% compared with total nucleosome monomers[31]) served as the histone source. As the chromatin was not digested with RNase before isolation of nucleosomes, considerable amounts of snRNA presumably complexed in snRNP[32] were also present (Chr. Elsner, personal communication).

A suitable DNA fragment containing a $tRNA_2^{Lys}$ gene was subcloned from cloned nucleosome tetramer DNA[17]. The orientation of the fragment with respect to restriction sites of the pBR322 plasmid vector and the position of nucleosomes on the corresponding DNA in chicken embryo chromatin are shown in Fig. 1*a*. To generate ssDNA stretches of defined length and orientation, a 30-bp fragment of radioactively labelled palindromic DNA was ligated to the gene sequence and the DNA upstream and downstream from the ligation site was differentially digested with exonuclease III. The manipulated DNA was ligated to an acceptor plasmid and the ends of the single-stranded stretches were annealed via the 30-bp palindromic DNA. In the ionic conditions used in subsequent experiments, the annealed DNA (sssDNA) was stable up to 40 °C. Oligomerization during ligation reactions was minimized and all recombined fragments had the same orientation within the acceptor plasmid. Moreover, a radioactive label was present at a defined site. The time course of the exonuclease III reaction was monitored by observing the disappearance of restriction sites, when the respective DNA base sequence became single-stranded. Several such sites were spaced in roughly 20-bp intervals or multiples thereof near the 30-bp palindromic DNA. As one restriction site always completely disappeared before the following one began to fade, the variation in length of the single strands at a given incubation time was <20 bp.

For the site-directed assembly process, the soluble nucleoprotein fraction, the histone source and ssDNA were combined to give a highly concentrated solution. The amount of protein provided by the nucleoprotein fraction and the quantity of nucleosomes in the histone source were each equivalent to $\sim 5\times 10^7$ nuclei. Compared with the total nuclear volume, both constituents were only 10-fold diluted; 10^{-10} mol of nucleosomes and $<5\times 10^{-13}$ mol of ssDNA were used in a standard assay. During incubation in the presence of the four deoxynucleoside triphosphates and ATP, double-stranded DNA was synthesized from ssDNA and concomitantly nucleosomes were assembled. The positions of the assembled nucleosomes were determined accurately using three complementary methods[33-35] (Fig. 1*b*). The nucleosome position map reveals that the middle axis of one nucleosome always coincided roughly with the midpoint of the ssDNA stretch in the ssDNA used. Single-stranded stretches longer than 120 nucleotides were degraded during incubation and if <40 nucleotides were single-stranded, the assembled nucleosomes associated randomly with respect to DNA base sequence. These data indicate that the stretch of ssDNA itself and/or a protein temporarily bound at about that position direct one nucleosome to a defined site.

tRNA gene transcription within randomly positioned nucleosomes

The soluble ribonucleoprotein fraction of the assembly extract should provide at least 0.02 units per assay of DNA-dependent RNA polymerase III (1 unit incorporates 1 nmol UMP into RNA in 10 min using calf thymus DNA as template[30]), but no such activity was detected in a complete assembly extract, and partially purified RNA polymerase III was soon inactivated when added to the extract. Therefore, after nucleosome assembly we removed most of the components of the assembly extract by brief ultracentrifugation. This did not yield a pure preparation of assembled nucleosomes but the added RNA polymerase III was now active. The concentrations of protein and DNA in the transcription assay were over one order of magnitude lower than in the assembly extract whereas the final concentrations of buffer components were the same. Higher Mg^{2+} concentrations led to precipitation of the assembled nucleosomes, and Mn^{2+} as the divalent cation did not allow for specific transcription. The amounts of RNA polymerase III added provided enzyme saturation.

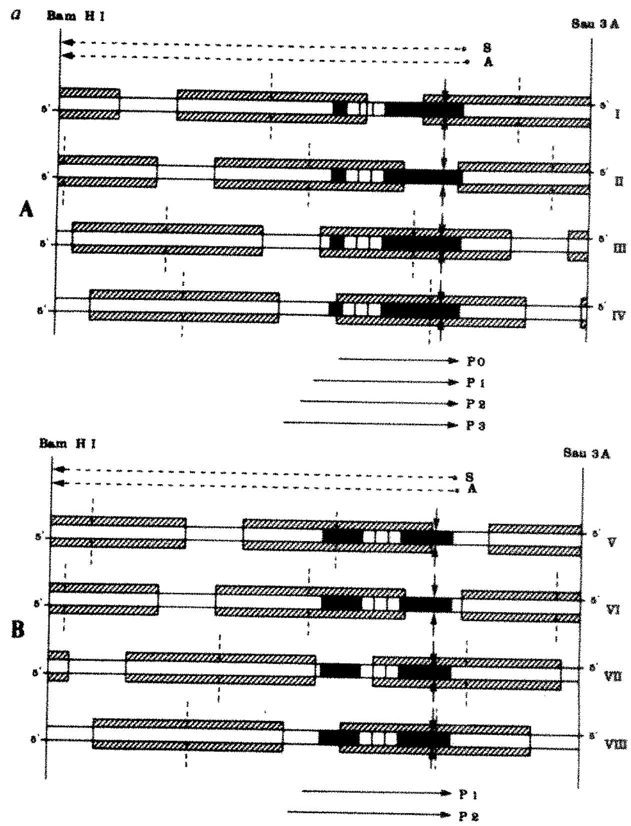
To evaluate transcription experiments, the reference used was transcription of randomly positioned nucleosomes; these were obtained by incubating double-stranded circular DNA with the reconstitution extract. The DNA either had no palindromic insertion or the 30-bp palindromic DNA was inserted into restriction sites for *Alu*I (type A insertion) and *Hin*fI (type B insertion), respectively (Fig. 2*a*). These templates within randomly positioned nucleosomes were transcribed by the added RNA polymerase III into several discretely sized RNA molecules (Fig. 2*b*) but there was no apparent mature tRNA synthesized. Nevertheless, several tRNA precursor samples were identified by S_1 nuclease hybridization mapping[36] (Fig. 2*c*).

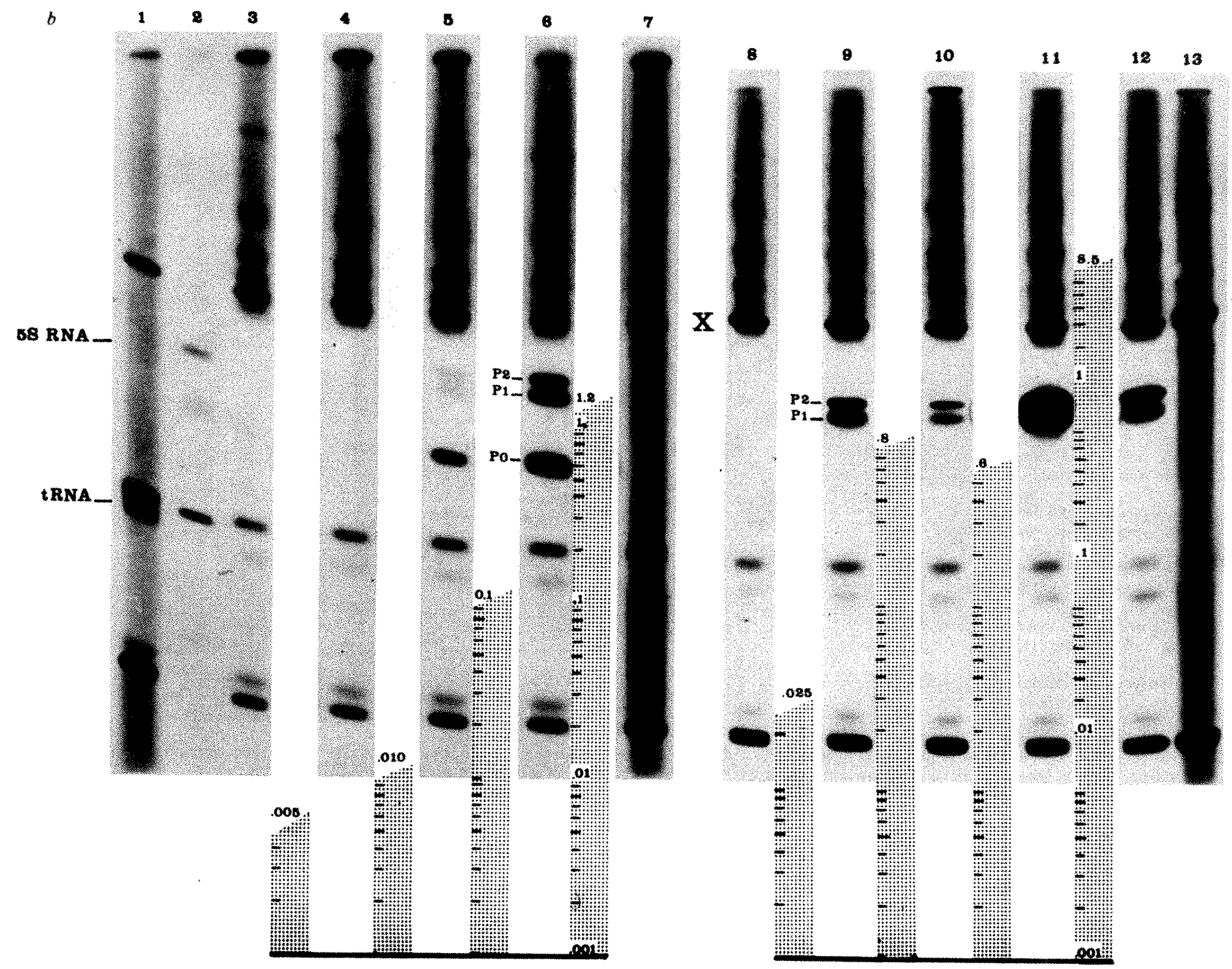
DNA lacking a palindromic insertion was transcribed to give three kinds of tRNA precursors denoted P1, P2 and P3; these amounted to 22.5% of total RNA synthesis, at a ratio of 3 : 6 : 1, respectively. Efficiency of transcription was low, only 1 in 10 tRNA genes being transcribed in 100 min. The precursors each began at the first guanosine residue of the sequence $AGGC^{A}_{T}$, which is repeated 10 (P1), 21 (P2) and 33 (P3) bp in front of the tRNA structural gene. Assay of randomly positioned nucleosomes containing the type A insertion revealed that P1 and P2 were 30 bp longer, P3 was not detected, and another precursor (P0) was predominant (the ratio P0/P1/P2 was 6 : 2 : 2). P0 starts within the tRNA at the 5′ end of the 30-bp palindromic insertion which gives in this position the sequence AGCCA; again the guanosine nucleotide is the initiation site. The amount of precursors synthesized was 9% of the total RNA synthesis; ~4% of the tRNA genes were transcribed in 100 min. Type B insertion DNA within randomly positioned nucleosomes was transcribed with the same efficiency as if no insert were present. P0 could not be detected and P2 was again the preferred precursor. Isolated DNA alone of any of the three types was not faithfully transcribed by RNA polymerase III.

tRNA gene transcription depends on nucleosome position

A strikingly different picture was obtained when the $tRNA_2^{Lys}$ gene was transcribed within site-directed nucleosomes. Four nucleosome positions were assayed which differed with respect to the sequence TTCGA. The four template situations were assembled using sssDNA and type A as well as type B insertions (Fig. 3*a*). TTCGA was either located near the border of a nucleosome (templates I and V); in the linker DNA (II and VI); ~20 bp from the nucleosome middle axis (III and VII); or very close to the middle axis (IV and VIII). Transcription of tRNA precursors was heavily suppressed in templates I and V (Fig. 3*b*) although total RNA synthesis corresponded to randomly positioned nucleosomes. Only 0.5% (I) and 2.5% (V) of the tRNA genes were transcribed. In contrast, when TTCGA was located close to the middle axis of a nucleosome (IV and VIII), transcription of tRNA precursors showed dramatic increases of 30-fold for template IV and 85-fold for template VIII, compared with corresponding control experiments using randomly positioned nucleosomes. This is equivalent to one and eight rounds of transcription, respectively, for each tRNA gene. The predominant precursors were P0 for type A insertion DNA (IV) and P2 when type B insertion was used to construct ssDNA. When the nucleosome middle axis was shifted only 20 bp to the right or left of the TTCGA sequence (experiments III and VII) the transcription of tRNA precursors decreased by one order of magnitude (this in turn demonstrates the accuracy of the nucleosome position analysis). Location of TTCGA in the nucleosome linker DNA yielded elevated precursor transcription with the type B insertion only. The amounts of tRNA precursors transcribed were varied over two orders of magnitude by site-directed assembly of nucleosomes. The $tRNA_2^{Lys}$

Fig. 3 *a*, Scheme of templates used for transcription: I–VIII. A, type A and B, type B insertion of 30-bp palindromic DNA; broken arrows labelled S and A, hybridization probes for sense and anti-sense transcription mapping, respectively; arrows labelled P0, P1, P2 and P3 represent mapping of tRNA precursors; vertical broken line, nucleosome middle axis; vertical arrows indicate location of TTCGA sequence. *b*, Transcription of the $tRNA_2^{Lys}$ gene within site-directed nucleosomes. 1, Assembly extract minus sssDNA; 2, chicken embryo $tRNA_2^{Lys}$ and *E. coli* 5S RNA, respectively; 3, expt I; 4, expt II; 5, expt III; 6, expt IV; 7, isolated DNA from expt IV; 8, expt V; 9, expt VI; 10, expt VII; 11, expt VIII; 12, expt VIII, $\frac{1}{10}$ of sample applied to electrophoresis; 13, isolated DNA from expt VIII. tRNA gene transcription was quantified as for Fig. 2*b*. A sheet of Whatman 3 MM paper was placed between the gel and film covering the area below band X for 2–6 and 8–12 (see also slots 6–8 in Fig. 2*b*). This allows better photographic documentation of the intensity range of tRNA precursor bands by improving the signal-to-background ratio in the relevant area of the gel.





gene investigated was most efficiently transcribed when the sequence TTCGA was associated with the nucleosome middle axis.

To validate these interpretations, we performed two control experiments (Fig. 4). First, after transcription, an aliquot of the transcription assay was routinely digested with restriction endonuclease *Bam*HI. The enzyme should cause cleavage if its site is located in the nucleosome linker DNA whereas cleavage should be suppressed if the site is associated with the nucleosome core (see Fig. 3*a*). DNA was purified and then restricted with *Sph*I, which cleaved once in the acceptor plasmid downstream from the chicken embryo DNA (see Fig. 1*a*). Positive *Bam*HI restriction in assembled nucleosomes yielded a radioactively labelled 591-bp fragment after subsequent *Sph*I restriction. In the case of suppressed cleavage, the circular DNA was just linearized by *Sph*I restriction.

Another aliquot of the transcription assay was digested extensively with *Staphylococcus* nuclease to analyse the amount of radioactive label (residing in the 30-bp palindromic DNA) that could be recovered in the nucleosome core DNA. Autoradiographs of the respective fragments separated by agarose gel electrophoresis are shown in Fig. 4*a*. Obviously, *Bam*HI cleaves template VIII but not template V. The strong band of radioactively labelled nucleosome monomer DNA indicates that in both templates the 30-bp palindromic DNA is associated with the nucleosome core. Taken together, these constitute good evidence that nucleosomes are still in position after transcription. Liquid scintillation counting of the excised agarose gel bands revealed that at least 90% of the nucleosomal templates used belong to the depicted population (Fig. 3*a*). If an unrecognized fraction representing <10% of the material, although having a pronounced effect, were the template, the efficiency of transcription would increase to unrealistic values[37]. We also consider it unlikely that small amounts of ssDNA stretches, persisting after site-directed nucleosome assembly, were the active templates. Templates VI and VII were transcribed with equal efficiency although the coding strand of VI and the noncoding strand of VII, respectively, had to be rendered single-stranded for site-directed nucleosome assembly.

The second control experiment was designed to elucidate further the proteins forming the template. To do this, quantitative amounts of nucleosomes were isolated from the template region. Although several restriction sites seemed suitable to excise the nucleosomal templates by cutting within the respective upstream and downstream nucleosome linker DNAs, only *Bam*HI was active in the transcription assay. Therefore, a subclone was constructed containing an additional *Bam*HI site instead of the downstream *Sau*3A site. Type VIII templates were assembled, incubated for a short period in transcription assays and the excised dimer nucleosomes were purified after *Bam*HI digestion. The yield of protein from 50 such assays (each containing 1 μg of sssDNA) was in the range expected for the dimer nucleosomes (~3–5 μg). SDS-polyacrylamide gel electrophoresis clearly showed (Fig. 4*b*) that the four nucleosome core histones are the predominantly extracted protein fractions. If other proteins were present in molar concentrations of only half those of the core histones, they would be detected on the

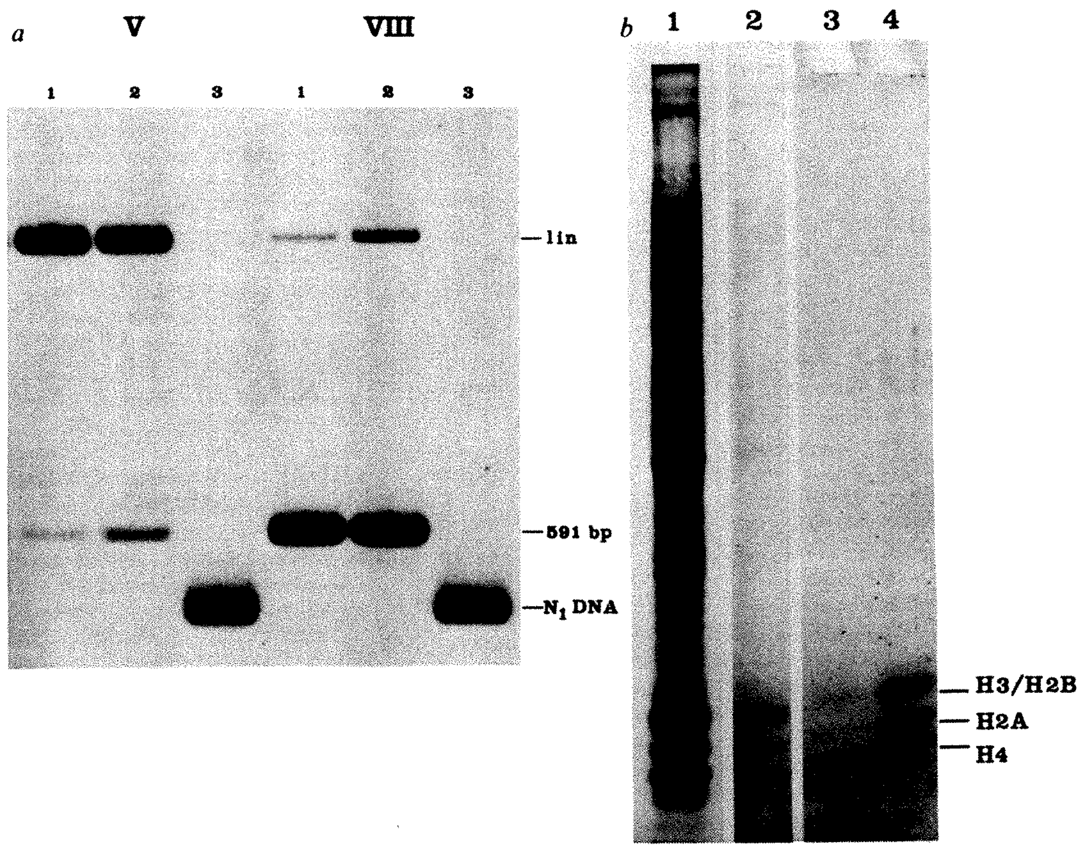


Fig. 4 *a*, Screening of nucleosome positions after tRNA gene transcription. V and VIII refer to the respective nucleosomal templates (see Fig. 3*a*). 1, 10 μl of transcription assay restricted with *Bam*HI; 2, 20 μl of transcription assay restricted with *Bam*HI; 3, 10 μl of transcription assay digested with staphylococcal nuclease. Lin, linearized DNA; N1 DNA, nucleosome monomer DNA. For *Bam*HI restriction, aliquots of the transcription assay were diluted to 50 μl with 150 mM NaCl, 10 mM $MgCl_2$, 5 mM 2-mercaptoethanol, 10 mM Tris-HCl *p*H 7.9. *Bam*HI (10 U) was added and the suspension incubated at 37 °C for 30 min with brief vortexing every 5 min. Nucleic acids were deproteinized and restricted with *Sph*I. For *Staphylococcus* nuclease digestion, aliquots of the transcription assay were diluted to 50 μl (equivalent to 5×10^7 nuclei ml^{-1}) with 0.25 mM EDTA, 20 mM Tris *p*H 7.9, and preincubated at 37 °C for 1 min. 1 μl of 50 mM $CaCl_2$ and 1 μl (1 U) of *Staphylococcus* nuclease were added and incubation continued for 20 min. Nucleic acids were deproteinized and separated on a 1% agarose gel. The gel was dried using DEAE-paper and autoradiographed. Polyacrylamide gel electrophoresis of proteins extracted from the transcription assay and from purified nucleosomal template VIII. 1, Proteins extracted from 15 μl (600 μg protein) of a complete nucleosome assembly extract (25 μl). 2, Total protein extracted from a transcription assay containing a type VIII nucleosomal template. 3, 1.5 μg of the nucleosome monomer preparation (0.7 μg of core histones) serving as the histone source in the site-directed nucleosome assembly system. 4, Total protein extracted from the purified dimer nucleosomes of type VIII nucleosomal templates. Type VIII nucleosomal templates were assembled using sssDNA with a second *Bam*HI site instead of the *Sau*3A site (see Fig. 1 legend). These templates were incubated in transcription assays at 37 °C for 15 min, the $MgCl_2$ concentration was adjusted to 10 mM, and the suspension digested with 100 U of *Bam*HI at 37 °C for 30 min with brief vortexing every 5 min. Fifty such assays were pooled and dialysed extensively against 20 mM KCl, 1 mM EDTA, 5 mM Tris-HCl *p*H 7.9. Samples (1 ml) were layered on to 6–30% (w/w) linear sucrose gradients and centrifuged in a SW41 rotor at 150,000*g* for 15 h. Fractions containing the dimer nucleosomes of interest (identified by the radioactive label residing in the 30-bp palindromic DNA) were recovered, extensively dialysed against H_2O and lyophilized. Proteins were extracted by incubation in 2% (w/v) SDS, 3% (v/v) 2-mercaptoethanol, 0.1% (w/v) phenylmethylsulphonyl fluoride, 10% (v/v) glycerol at 95 °C for 5 min. Proteins were separated on a 5–20% (w/v) polyacrylamide linear gradient gel using as buffer 50 mM sodium phosphate, 0.1% (w/v) SDS. The gel was stained with Coomassie brilliant blue, R-250.

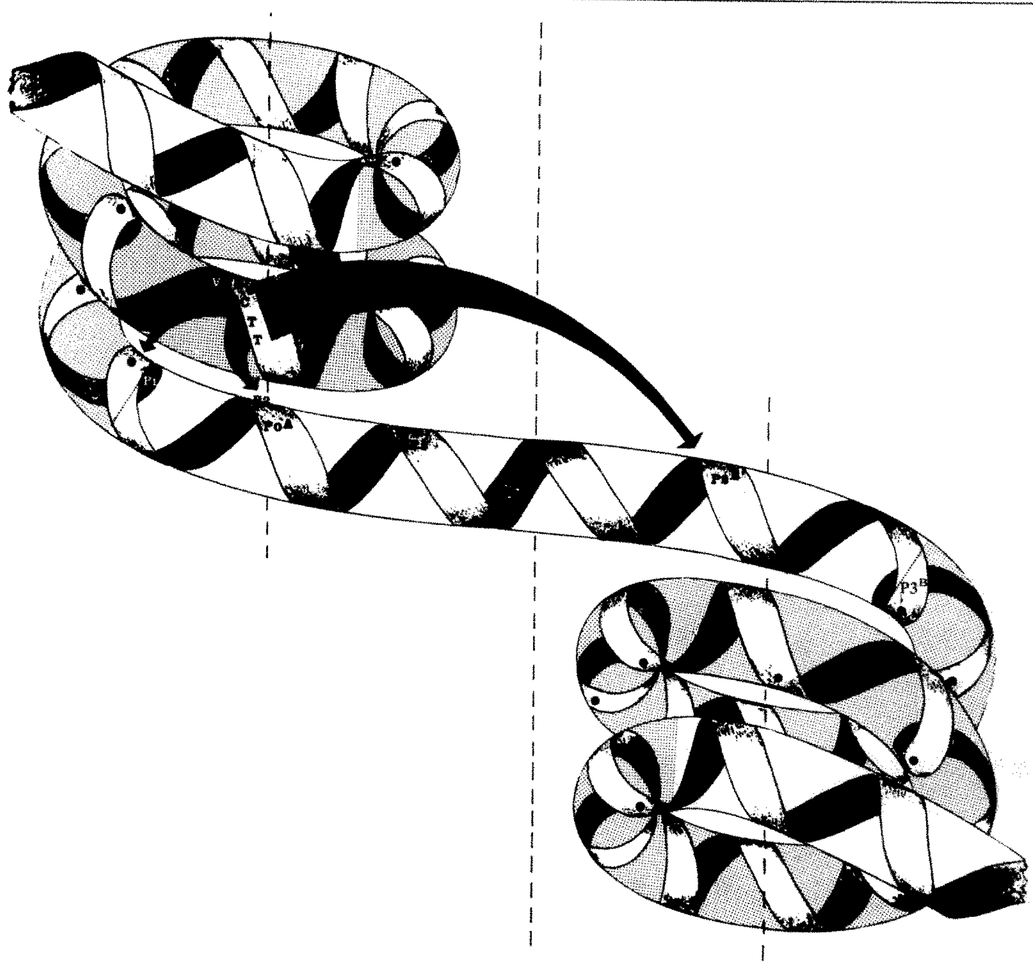


Fig. 5 Functional tRNA gene promoter structure. DNA of the nucleosome core is shown against a shaded background. P1, P2 and P3 represent initiation sites used *in vivo*; $P0^A$, $P1^B$ and $P2^B$ are initiation sites in assembled nucleosomes containing type A and type B insertion DNA, respectively. $P3^B$ is not used for initiation.

gel. Thus, the protection against nucleases which produced the corresponding nucleosome position maps (Figs 1*b* and 3*a*) could not be effected by proteins other than histones.

Nevertheless, additional proteins are required to promote efficient transcription of the tRNA gene within the assembled nucleosomes. The transcription assay contained a heterogeneous mixture of proteins although it was partially purified from excess reconstitution extract and histone source to allow for RNA polymerase III activity (Fig. 4*b*). However, so far any attempt to purify the nucleosomal template further drastically reduced transcription efficiency. When the RNA polymerase III preparation was added to a finally purified template such as the one used for the protein analysis above, faithful tRNA gene transcription was not observed. We are now investigating which components of the assembly extract and transcription assay are ultimately required using immunological methods.

Conclusions

The construction of sssDNA requires the insertion of a 30-bp palindromic DNA fragment into the tRNA gene. The influence of this insertion on transcription efficiency had to be elucidated before the effect of different nucleosome positions could be interpreted. In this context we have found that insertion of unrelated DNA into the 5′-end region of the tRNA gene (type A insertion) decreases transcription of the gene, whereas the corresponding insertion is tolerated near the middle of the gene (type B insertion). These results are consistent with recent data on consequences of sequence substitutions within a $tRNA_1^{Met}$ gene[25].

The three $tRNA_2^{Lys}$ precursors observed were not processed to mature tRNA in the *in vitro* system used. We cannot exclude, however, the possibility that precursor P3 alone represents initiation of transcription and that P2 and P1 are equivalents of processing steps[38–40].

The model[41] in Fig. 5 shows the three-dimensional aspects of a functional tRNA gene promoter. The *in vivo* location of the TTCGA sequence within nucleosomes for chicken embryo tRNA genes is identical to the most efficiently transcribed template of nucleosomes assembled *in vitro*. In both cases the TTCGA sequence—the region of symmetry for a 5-bp complementary palindrome—is associated with a region of symmetry of the nucleosome core, that is, the middle axis in two-dimensional schemes. We assume that this structure is primarily recognized by RNA polymerase III, as a deviation from the coincidence of TTCGA with the nucleosome middle axis decreased transcription of the gene. Location of TTCGA near the middle of the nucleosome linker DNA may have an equivalent function (see template VI) although this situation was not detected *in vivo*. The preferred *in vivo* initiation site, P2, and $P0^A$, the incorrect but preferred initiation site with type A insertion DNA, are located near to the TTCGA sequence. However, the distance to TTCGA is not the main selection principle, as with type B insertion DNA transcription starts predominantly at $P2^B$ and not at the closer $P1^B$ site. The orientation of $P2^B$ with respect to the next nucleosome, which is similar to that of P2 in the first nucleosome, may be the reason for its selection. As we observed no initiation at the hypothesized $P3^B$ site, $P2^B$ may also represent the farthest possible distance from TTCGA. The distance between P1 and $P2^B$, where initiation sites should principally be located, provides sufficient space to tolerate introns in tRNA genes[39,40,42]. (In a recent paper[17] a figure shows two cloned $tRNA^{Lys}$ genes containing introns. From the established sequence we now consider these introns to be artefacts because they are duplications of acceptor stem and dihydrouridine stem sequences. They probably originate from the tRNA gene enrichment procedure which involved repair synthesis of double-stranded DNA by *E. coli* DNA polymerase I. Moreover, we were unable to isolate $tRNA^{Lys}$ genes containing introns from a chicken

DNA library (gift of J. D. Engel), although genes corresponding to the one described in the present article were detected.)

We have shown that the functional state of a tRNA gene can be programmed by nucleosome positions. The conserved short DNA base sequence TTCGA which occurs too frequently in the large chicken genome to have any functional significance, is converted into an efficient promoter through a specific arrangement in the nucleosome.

We thank Dr H. Tiedemann for discussion, Chr. Elsner, S. Hense and Chr. Keitel for help in experimentation, and J. Staubach for the graphic design. This work was supported by the Deutsche Forschungsgemeinschaft through Sonderforschungsbereich 29.

Received 7 October 1981; accepted 26 March 1982.

1. Moyne, G., Katinka, M., Saragoshi, S., Chestier, A. & Yaniv, M. *Prog. Nucleic Acid Res. molec. Biol.* **26**, 151-167.
2. Weintraub, H. & Groudine, M. *Science* **193**, 848-856 (1976).
3. Stalder, J., Groudine, M., Dodgson, J. B., Engel, J. D. & Weintraub, H. *Cell* **19**, 973-980 (1980).
4. Bellard, M., Kuo, M. T., Dretzen, G. & Chambon, P. *Nucleic Acids Res.* **8**, 2737-2750 (1980).
5. Gariglio, P., Bellard, M. & Chambon, P. *Nucleic Acids Res.* **9**, 2589-2598 (1981).
6. Levy, A. & Noll, M. *Nature* **289**, 198-203 (1981).
7. Storb, U., Wilson, R., Selsing, E. & Walfield, A. *Biochemistry* **20**, 990-996 (1981).
8. Groudine, M. & Weintraub, H. *Cell* **24**, 393-401 (1981).
9. Wu, C. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1577-1580 (1981).
10. Levy, W. B. & Dixon, G. *Nucleic Acids Res.* **5**, 4155-4163 (1978).
11. Chahal, S. S., Matthews, H. R. & Bradbury, E. M. *Nature* **287**, 76-79 (1980).
12. Groudine, M., Eisenman, R. & Weintraub, H. *Nature* **292**, 311-317 (1981).
13. Grosschedl, R. & Birnstiel, M. L. *Proc. natn. Acad. Sci. U.S.A.* **77**, 7102-7106 (1980).
14. Mathis, D., Oudet, P. & Chambon, P. *Prog. Nucleic Acid Res. molec. Biol.* **24**, 2-49 (1980).
15. Zachau, H. G. & Igo-Kemenes, T. *Cell* **24**, 597-598 (1981).
16. Kornberg, R. *Nature* **292**, 579-580 (1981).
17. Wittig, B. & Wittig, S. *Cell* **18**, 1173-1183 (1979).
18. Ruiz-Carrillo, A., Jorcano, J. L., Eder, G. & Lurz, R. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3284-3288 (1979).
19. Philip, M., Jamaluddin, M., Sastry, R. V. R. & Chandra, S. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5178-5182 (1979).
20. Laskey, R. A. & Earnshaw, W. C. *Nature* **286**, 763-767 (1980).
21. Weissbach, A. *A. Rev. Biochem.* **46**, 25-47 (1977).
22. Palter, K. B., Foe, V. E. & Alberts, B. M. *Cell* **18**, 451-467 (1979).
23. Gould, H. J., Cowling, G. J., Harborne, N. R. & Allan, J. *Nucleic Acids Res.* **8**, 5255-5266 (1980).
24. Wittig, S. & Wittig, B. *Nucleic Acids Res.* (in the press).
25. Hofstetter, H., Kressmann, A. & Birnstiel, M. L. *Cell* **24**, 573-585 (1981).
26. Sakonju, S., Bogenhagen, D. F. & Brown, D. D. *Cell* **19**, 13-25 (1980).
27. Bogenhagen, D. F., Sakonju, S. & Brown, D. D. *Cell* **19**, 27-35 (1980).
28. Dykes, G., Bambara, R., Marians, K. & Wu, R. *Nucleic Acids Res.* **2**, 327-345 (1975).
29. Wittig, S. & Wittig, B. *Hoppe-Seyler's Z. physiol. Chem.* **362**, 379-387 (1981).
30. Wittig, B. & Wittig, S. *Biochim. biophys. Acta* **520**, 596-611 (1978).
31. Wittig, B. & Wittig, S. *Nucleic Acids Res.* **4**, 3901-3917 (1977).
32. Lerner, M. R. & Steitz, J. A. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5495-5499 (1979).
33. Igo-Kemenes, T., Omori, A. & Zachau, H. G. *Nucleic Acids Res.* **8**, 5377-5389 (1980).
34. Dingwall, C., Lomonosoff, G. P. & Laskey, R. A. *Nucleic Acids Res.* **9**, 2659-2673 (1981).
35. Hörz, W. & Altenburger, W. *Nucleic Acids Res.* **9**, 2643-2657 (1981).
36. Sharp, P. A., Berk, A. J. & Berget, S. M. *Meth. Enzym.* **65**, 750-758 (1980).
37. Parker, C. S. & Roeder, R. G. *Proc. natn. Acad. Sci. U.S.A.* **74**, 44-48 (1977).
38. Birkenmeyer, E. H., Brown, D. D. & Jordan, E. *Cell* **15**, 1077-1086 (1978).
39. Ogden, R. C. *et al. Cell* **17**, 399-406 (1979).
40. Mao, J., Schmidt, O. & Söll, D. *Cell* **21**, 509-516 (1980).
41. Finch, J. I. *et al. Nature* **269**, 29-36 (1977).
42. Robinson, R. R. & Davidson, N. *Cell* **23**, 251-259 (1981).
43. Vogelstein, B. & Gillespie, D. *Proc. natn. Acad. Sci. U.S.A.* **76**, 615-619 (1979).

LETTERS TO NATURE

NGC5128—a galaxy with a recently formed disk

Michel Marcelin & Jacques Boulesteix

Observatoire de Marseille, 2 place Le Verrier, 13248 Marseille Cedex 4, France

Georges Courtes & Bruno Milliard

Laboratoire d'Astronomie Spatiale, Les Trois Lucs, 13012 Marseille, France

The galaxy NGC5128, better known as the powerful radio source Centaurus A, is a puzzling object appearing as an elliptical galaxy crossed by a dense lane of dust and gas. Its structure and evolution are not well understood, although this object is quite near (we shall adopt 5 Mpc following Burbidge and Burbidge[1]), bright, and has been widely observed in the past few years. We have obtained the first complete velocity field of the ionized gas in NGC5128, with a 10-km s^{-1} precision, from Perot–Fabry interferometry with the ESO 3.6-m telescope. It shows that the gas is rotating in a thick disk (~1 kpc thickness), without any evidence of the perturbations expected from the encounter of an elliptical and a spiral galaxy. Our observations favour the hypothesis of the evolution of a spiral galaxy in which the formation of the disk is very late compared with the bulge.

The velocities in this object were first studied by Burbidge and Burbidge[1] with a slit spectrograph and a 2-m telescope. It was then observed by Sersic[2] with a 1.5-m telescope, using slit spectrography (mainly for the kinematics of the elliptical component through absorption lines) and Perot–Fabry interferometry (in Hα line) for the kinematics of the gas. Eventually Graham[3] made a good study of the rotation of the gas and of the motions in the elliptical component with a slit spectrograph attached at the CTIO 4-m telescope.

The more recent photometric study by Dufour *et al.*[4] gives mainly photometric parameters of the elliptical component; the gas being more difficult to study because of the high inclination and thickness of the disk, and because of the great quantity of dust.

Our study of the gas in NGC5128 has been made through a narrow bandwidth filter (15 Å FWHM), centred at 6,575 Å, isolating the Hα emission line of the ionized gas. Comte and Georgelin took Hα plates of the galaxy with a focal reducer and an image tube attached at the ESO 3.6-m telescope. Four interferograms of the object taken with the same device and a Perot–Fabry interferometer (see Table 1) gave interesting information on the morphology of the ionized gas and particularly the kinematics of the gas. They provide the first complete velocity field of the ionized gas in this object.

A preliminary analysis[5] suggested that the gas is approximately distributed in a disk inclined at ~75° with respect to the line of sight; the temperature of this ionized gas being quite normal as shown by the relatively thin interference lines (<1 Å wide) which indicate no peculiar turbulence[6].

Such an observation suggests that NGC5128 is not the result of a collision in which a spiral galaxy would have merged into an elliptical galaxy as has been often suggested. The following complete analysis of the velocity field of NGC5128 confirms this opinion.

Another interesting detail is the detection of three bubble-like H II regions[5] of ~200 pc in diameter (assuming that the galaxy is at 5 Mpc). This is quite comparable with bubble regions observed in nearby spiral galaxies such as M31 (ref. 7), M33 (ref. 8) or NGC925 (ref. 9).

Concerning the morphology of the galaxy itself, we obtained a striking image by viewing simultaneously on a screen a Hα plate (projected through a red filter) and a visible plate (projected through a green filter). The two components of the galaxy are quite distinct: the old elliptical component appears as a yellow ball, while the surrounding gas and dust show up vividly in red and black. The plates used for this composite image are I 1864 (40-m in exposure) and I 1886 (30 s exposure), see Table 1. Due to our very narrow Hα interference filter (15 Å FWHM) and our fast aperture ratio, $f/2$ with a 3.6-m telescope, the H II regions are much clearer than on the colour photograph obtained by Dufour *et al.*[4,10] with a larger Hα filter (54 Å FWHM) but with a slow aperture ratio, $f/10$ with a 1-m telescope. This is especially true a long way from the nucleus where

0028-0836/82/180038—05$01.00

Table 1 Plate material

Plate no.	Date	Exposure (min)	Emulsion	Filter	Focal reducer	Perot–Fabry (interference order)	Quality	No. of velocity points per interferogram
I 1864	12.02.78	10*, 40*	IIIaJ (B), N_2	Hα (15 Å)	f:1.9	No	Very good	
I 1881	4.04.78	150	098.02 (B), N_2	Hα (15 Å)	f:0.95	1051	Good	135
I 1886	5.04.78	0.2*, 0.5*, 1.5*	IIIaJ (B), N_2	Orange (wide)	f:1.9	No	Good	
I 1887	5.04.78	90*	IIIaJ (B), N_2	Hα (15 Å)	f:1.9	1051	Very good	288
I 1888	5.04.78	60*	IIIaJ (B), N_2	Hα (15 Å)	f:1.9	1051	Good	241
I 1889	5.04.78	60*	IIIaJ (B), N_2	Hα (15 Å)	f:1.9	1051	Good	214

* With RCA image intensifier.

one can see nothing but dust on their photograph (image tubes were used by both teams).

The measurement of the four interferograms gave about 900 velocity points which were averaged in 210 squares of 15 arc s a side on the sky (see Fig. 1). The aim of averaging the velocity points is to reduce the internal dispersion of the measurements and to facilitate the drawing of the isovelocity lines. In each square we computed the dispersion of the measurement (not taking into account the dispersion inside each single interferogram), which gives an idea of the expected precision for the value of the radial velocity. We then computed the statistical error, for each square, defined as $s/\sqrt{n}$ where n is the number of interferograms in the considered square. The average of the statistical error indicates a mean precision of 11 km s^{-1} on the whole velocity field. This allows us to draw the isovelocity lines with a 30 km s^{-1} step.

The general design of the velocity field (see Fig. 2) is quite comparable with the velocity field of spiral galaxies such as M83 (NGC5236) (ref. 11), NGC7793 (ref. 12), or NGC253 (ref. 13) (although there is evidence for non-axisymmetric motions in the central parts of this galaxy, possibly due to a weak bar). This supports the idea that the ionized gas is laying in a disk, as already suggested from the elliptical shape of the distribution of the H II regions[3] and from the colour analysis[4]. The superimposition of all of our interference fringes on the same picture also indicates that the disk is inclined to roughly 75° with respect to the line of sight. This last method is very powerful for the Hα detection because of the high contrast it offers[14], about 1,000 times the contrast of a 250 Å wide colour filter. Nevertheless, both morphological and kinematical points of view rule out the possibility of a cylindrical structure as suggested by Rodgers and Harding[15]. Furthermore, there is no evidence of strong non-circular motions in our velocity field. However, near the minor axis the isovelocity lines are not exactly perpendicular to the major axis, their inclination suggesting some accretion (if one considers that the northern side of the disk is nearer than the southern one, as clearly shown by the dust lane appearance).

The most noticeable perturbation in our velocity field is at ~1 arc min west of the nucleus and can be explained by a strong thickness of the gaseous disk which would be mainly concentrated in a thick belt surrounding the elliptical component. The

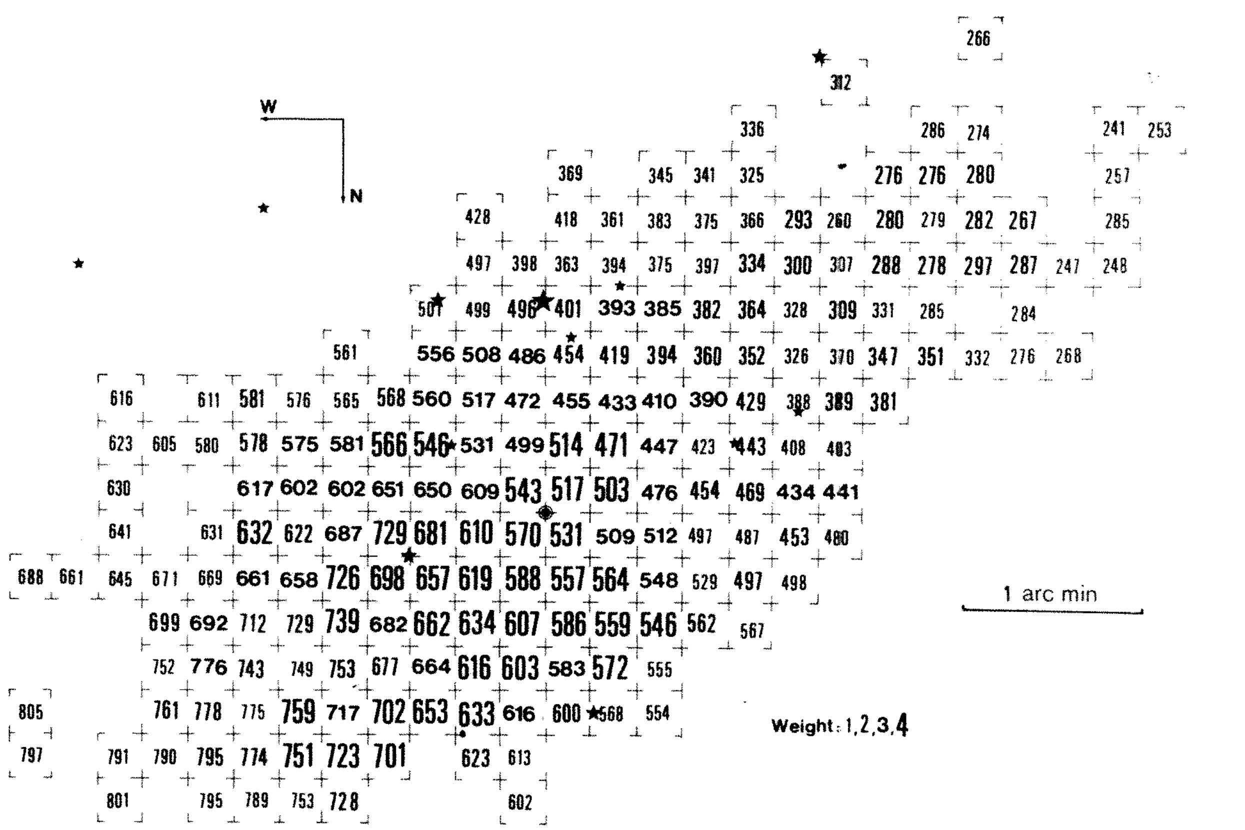


Fig. 1 Velocity measurements in NGC5128 averaged in squares of 15 arc s side. Four different sizes of figures are used, depending on how many interferograms (1, 2, 3 or 4) are averaged in a given square. The position of the nucleus is indicated (dotted circle) also with reference stars.

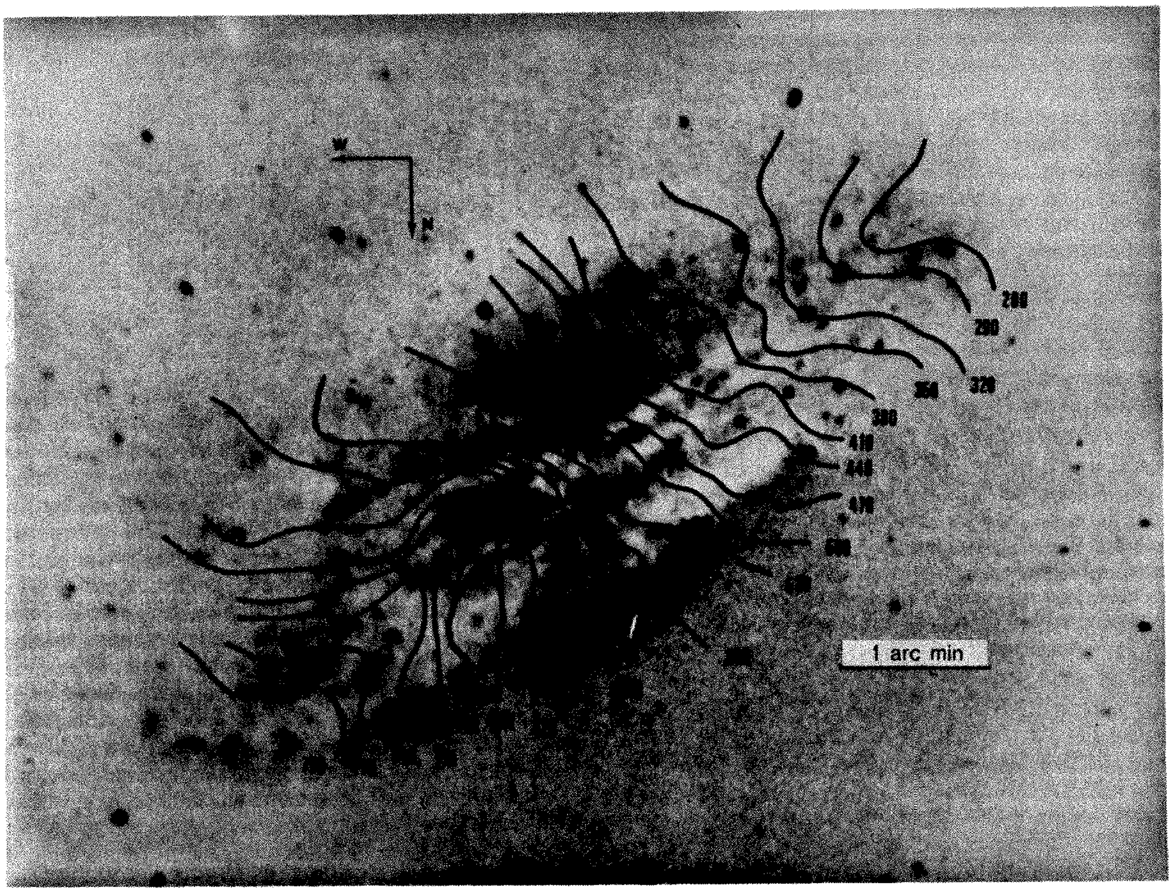


Fig. 2 Isovelocity lines superimposed on a Hα photograph of NGC5128. These lines are eye-estimates from Fig. 1, taking into account the weight of the data.

location of the maximum velocity gradient in this zone indicates that we are looking at gas $\sim$1 kpc above the mean plane of the disk, this gas masking the underlying disk. Such an extension of matter out of the mean galactic plane is confirmed by the morphology of the galaxy where one can see dust high above this plane silhouetted on the elliptical component. This would also explain the velocity anomalies stressed by Möllenhoff[16]. However, we could not confirm the superposition on the line of sight of objects with different radial velocities mentioned by some observers[9,15].

Another interesting detail is the straight chain of dust and H II regions stretching along 2 arc min from the nucleus towards the south-east. This chain also seems to be out of the mean plane (because it is silhouetted in front of the elliptical component while the nucleus is seen slightly under it when looking at the disk from above, with north down) and may be connected to the western part mentioned above. When studying the rotation curve of the disk, the assumption that it is flat would diminish the velocity gradient in the central part of the galaxy because the gas along the major axis is masked by gas above the plane. This gas being out of the plane of the disk, and away from the major axis, indicates slower radial velocities than expected on this line of sight and hence apparently slower rotation velocities. This remark will be helpful when analysing the rotation curve in the first 2 arc min from the nucleus.

Taking into account the squares lying within 33° (in the sky) of the major axis we obtain the rotation curve of Fig. 3 where we also plotted Graham's points[3]. The agreement is fairly good between the two sets of data. The parameters giving the least dispersion for our points are close to Graham's parameters[3]. From these data we have found: inclination of the plane of the galaxy with respect to the plane of the sky, $i = 73° \pm 3°$; position angle (PA) of major axis $= 127° \pm 5°$; systemic velocity $V_S = 545$ km s$^{-1} \pm 5$ km s^{-1} (while Graham[3] found, respectively, $i = 73°$, PA $= 125°$ and $V_S = 548$ km s^{-1}).

These last values of inclination and position angle of the major axis agree well with previous works[1,2,4,16]. The main problem concerns the systemic velocity for which Burbidge and Burbidge[1] found 605 km s^{-1}, Sersic[2] 462 km s^{-1}, Kunkel and Bradt[17] 508 km s^{-1}, Möllenhoff[16] 602 km s^{-1}. Our result, 545 km s^{-1}, is in very good agreement with the most recent papers of Graham[3], 548 km s^{-1}, or Whiteoak and Gardner[18] who find 551 km s^{-1} from H I observations.

Up to 4 arc min from the nucleus the symmetry of the two quadrants of the rotation curve is pretty good. Beyond 4 arc min we have almost no information for the southeastern quadrant, although we observe (somewhat lessened) the same tendency as Graham. That is, we see an increase of the velocity gradient in the outer part of the disk, which is probably due to matter moving out of the mean plane. Between 4 and 7 arc min in the northwestern quadrant the velocity gradient seems, in contrast, to diminish rapidly. Despite the small number of points available there, it seem probable that the rotation curve reaches a maximum between 5 and 7 arc min. The dashed line of Fig. 3 indicates a maximum rotation velocity $V_M = 352$ km s^{-1} at $r = 5.35$ arc min. This symmetric curve is a least square fitting polynomial (of the third degree) obtained from our points within 20° of the major axis to which we added the two outermost points given by Graham[3], also with our outermost point (this last point is in fact at 52° from the major axis but it is in very good agreement with Graham's outermost points[3]). Although their weight is rather poor and they may be out of the mean plane of the disk, it is the only way of extending our rotation curve beyond 4 arc min.

Following the method of Burbidge *et al.*[19] this least square curve indicates a mass $M = 0.9 \times 10^{11} M_\odot$ to $1.0 \times 10^{11} M_\odot$

within 4 arc min and a mass $M = 1.8 \times 10^{11} M_\odot$ to $2.0 \times 10^{11} M_\odot$ within 7 arc min, depending on the excentricity of the ellipsoïds used in the model. This value is smaller than the rough estimate $3 \times 10^{11} M_\odot$ given by Graham[3] from Kepler's law applied to the outermost points.

We have tried a more sophisticated method developed by Monnet and Simien[20] in which a model galaxy is built up from two components: an elliptical bulge and a flattened disk. This model shows that the bulge generally gives rise to a very steep velocity gradient near the nucleus, while our observed rotation curve surprisingly looks like a pure disk rotation curve, thus suggesting at once that we are dealing with a massive disk and a rather light bulge. The photometric parameters used for the spherical bulge were those given by Dufour *et al.*[4] (equivalent radius $r_e = 305$ arc s with a surface brightness $\sigma_B\ (r_e) =$ 22.94 mag arc s^{-2}, which gives a luminosity $I_e = 44.1 \mathscr{L}_\odot\ \mathrm{pc}^{-2}$). For the disk we used the photometric parameters given by Freeman[21] from Sersic's isophotes ($r_e = 336$ arc s, $I_e =$ $34 \mathscr{L}_\odot\ \mathrm{pc}^{-2}$), although, because of the great quantity of dust and because of the high inclination these parameters are not very reliable.

Assuming a constant $M/L_B = 5$ for the bulge, which is the value expected for such an elliptical component[20,22] we find a theoretical rotation curve for the bulge alone where the gradient near the nucleus is five times the observed gradient, with a maximum of more than 300 km s^{-1} near 1.5 arc min, before slowly decreasing. Comparison with Fig. 3 shows that the discrepancy between this model and the observations is much too high, even if we take into account the projection effect mentioned above, which would lessen the gradient in the inner part of the curve. Values of M/L_B between 1 and 3 for the bulge, comparable to what is found in globular clusters[23], are far more acceptable when comparing the model with the observations.

Eventually we found a satisfactory fit with an average $M/L_B \sim 2$ for the bulge and $M/L_B \sim 30$ for the disk, which is rather high but seems acceptable for a spiral galaxy[20]. From the total luminosity of each component we thus deduce a total mass of $1.1 \times 10^{11} M_\odot$ for the bulge, $8.2 \times 10^{11} M_\odot$ for the disk and $9.3 \times 10^{11} M_\odot$ for the galaxy as a whole. An intermediate solution with a varying M/L_B for the bulge (from 1 at the centre, up to 5 around 5 arc min) and a constant $M/L_B \sim 15$ for the disk could give a better fit, but it is not very realistic to assume a varying M/L_B for such an homogeneous stellar population. The mass would be then around $2 \times 10^{11} M_\odot$ for the bulge and $4 \times 10^{11} M_\odot$ for the disk, thus giving a total mass around $6 \times 10^{11} M_\odot$. The result remains much greater than our first estimate, based on the rotation curve, mainly because it relies on the outermost photometric data and hence takes into account a much greater part of the galaxy down to the sky background.

Note that, in any case, the disk component is found to be more massive than the elliptical component, while the bulge is much more luminous.

We will now try to review briefly some of the different hypotheses concerning the possible evolution of NGC5128.

(1) The encounter of an elliptical and a spiral galaxy is strongly suggested by the two clearly different populations (an old stellar population in a bright elliptical bulge and a young stellar population, rich in gas and dust, in a disk). The main problem is that there is no evidence for the encounter, such as strong erratic motions in the gas. Furthermore, there is no significant difference between the systemic velocity we found for the disk, $V_S = 545$ km s$^{-1} \pm 5$ km s^{-1}, and the velocity found by Graham[3] for the elliptical component, 536 km s$^{-1} \pm 30$ km s^{-1}. Also, there is no trace of the nucleus supposed to have belonged to the spiral galaxy as Dufour *et al.*[4] emphasize. Then it is surprising that the main axis of the slightly prolate bulge (PA = 35°)[4] and the rotation axis of the disk (PA = 37°) are the same, within 2°. At least there are many other comparable galaxies having a prolate spheroïd as a bulge, which suggest that they are a peculiar type of object rather than the result of an encounter. For example: NGC5363, NGC1947, Cyg A and PKS1934 − 63 (ref. 24), NGC2865 (ref. 25) (the Spindle galaxy) and NGC612 (ref. 26) which is a radiogalaxy quite similar to NGC5128 in both radio and optical wavelengths.

(2) There are fewer objections to accretion of a gas cloud by an elliptical galaxy than there are for the above hypothesis. Furthermore, as already mentioned by Graham, the huge dust lane on the eastern side of NGC5128 seems to indicate the direction of accretion of matter coming from outside of the disk plane. However, the accretion of a primeval gas cloud can hardly explain why there is so much dust and why the H II regions are as metal rich as normal H II regions of spiral galaxies[4]. Furthermore, this hypothesis would imply a peculiarly favourable accretion angle to make the main axis of the disk and bulge coincide as well. An interesting model of accretion, studied by Tubbs[27], supposes that the mass of the disk is negligible, which is far from being the case as we have seen above.

(3) We prefer the hypothesis of a normal spiral galaxy (with late formation of the disk) of SO or Sa type, composed of a bright bulge and a thick disk. This disk is still under formation, undergoing accretion of dust and gas and slowly settling down. Larson's theory[28] on the evolution of elliptical and spiral galaxies could help explaining such an object if one assumes

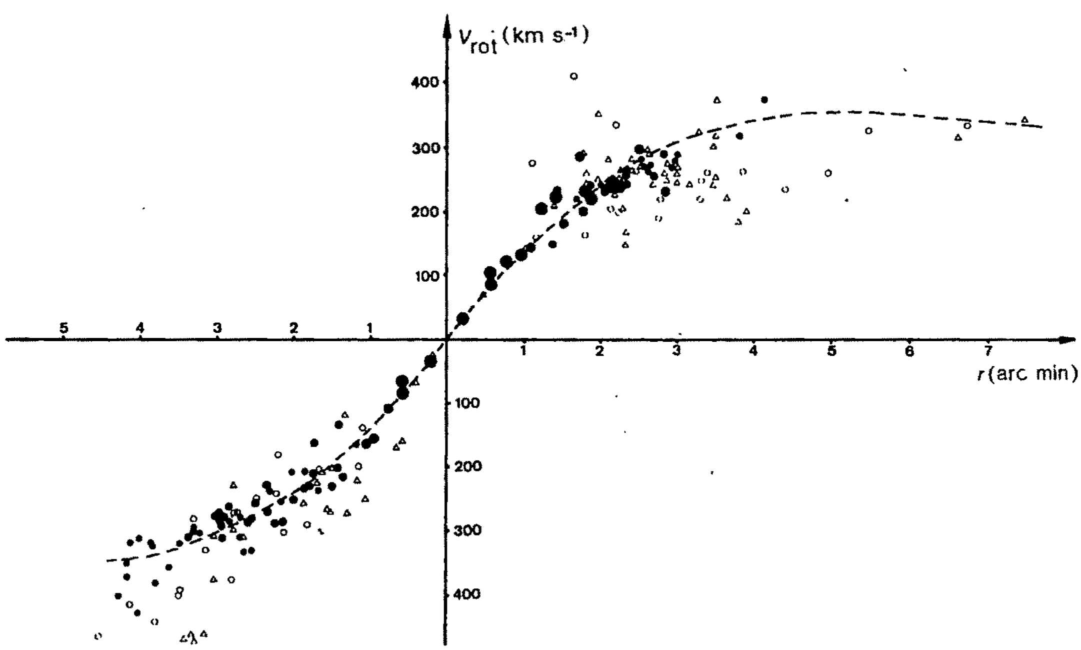


Fig. 3 Rotation curve of NGC5128. ●, Our velocity points within 20° from major axis which have different sizes, depending on their weight on Fig. 1; ○, our velocity points between 20° and 33° from major axis; △, Graham's velocity points. The dashed line is a least square fitting curve (see text).

that the protogalactic cloud was very large with a high density at its centre.

In the Larson's hypothesis[28] of a two-phase structure the dense clouds will rapidly form stars in a spheroïdal component (which may be prolate as showed by Gott and Thuan[29]), while the less dense intercloud gas will take a long time to settle to a disk and form stars. The continuing formation of this disk could explain the activity of NGC5128 and the associated powerful radiosource.

Received 1 February; accepted 6 April 1982.

1. Burbidge, E. M. & Burbidge, G. *Astrophys. J.* **129**, 271–281 (1959).
2. Sersic, J. L. *Nature* **224**, 253–254 (1969).
3. Graham, J. A. *Astrophys. J.* **232**, 60–73 (1979).
4. Dufour, R. J. *et al. Astr. J.* **84**, 284–301 (1979).
5. Marcelin, M. & Milliard, B. *Ann. Phys. Fr.* **4**, 139–146 (1979).
6. Boulesteix, J. & Courtès, G. *The Messenger* **14**, 2–3 (1978).
7. Pellet, A. *et al. Astr. Astrophys. Suppl.* **31**, 439–460 (1978).
8. Boulesteix, J., Courtès, G., Laval, A., Monnet, G. & Petit, H. *Astr. Astrophys.* **37**, 33–48 (1974).
9. Marcelin, M., Boulesteix, J. & Courtès, G. *Astr. Astrophys.* **108**, 134–140 (1982).
10. Dufour, R. J. & van den Bergh, S. *Sky Tel.* **56**, 389–395 (1978).
11. Comte, G. *Astr. Astrophys. Suppl.* **44**, 441–450 (1981).
12. Davoust, E. & de Vaucouleurs, G. *Astrophys. J.* **242**, 30–52 (1980).
13. Pence, W. D. *Astrophys. J.* **247**, 473–483 (1980).
14. Courtès, G. *Vistas Astr.* **14**, 81–161 (1972).
15. Rodgers, A. W. & Harding, P. *Astrophys. J. Lett.* **236**, L17–L21 (1980).
16. Möllenhoff, C. *Astr. Astrophys.* **93**, 248–254 (1981).
17. Kunkel, W. E. & Bradt, H. V. *Astrophys. J. Lett.* **170**, L7–L10 (1971).
18. Whiteoak, J. B. & Gardner, F. F. *Proc. astr. Soc. Austr.* **3**, 319–321 (1979).
19. Burbidge, E. M., Burbidge, G. R. & Prendergast, K. A. *Astrophys. J.* **130**, 739–748 (1959).
20. Monnet, G. & Simien, F. *Astr. Astrophys.* **56**, 173–180 (1977).
21. Freeman, K. C. *Astrophys. J.* **160**, 811–830 (1970).
22. Faber, S. M. & Jackson, R. E. *Astrophys. J.* **204**, 668–683 (1976).
23. Illingworth, G. *Astrophys. J.* **204**, 73–93 (1976).
24. Bertola, F. & Galletta, G. *Astrophys. J. Lett.* **226**, L115–L118 (1978).
25. Sandage, S. *The Hubble Atlas of Galaxies* (Carnegie Institution of Washington, 1961).
26. Ekers, R. D., Gross, W. M., Kotanyi, C. G. & Skellern, D. J. *Astr. Astrophys.* **69**, L21–L24 (1978).
27. Tubbs, A. D. *Astrophys. J.* **241**, 969–980 (1980).
28. Larson, R. B. *Mon. Not. R. astr. Soc.* **176**, 31–52 (1976).
29. Gott, J. R. & Thuan, T. X. *Astrophys. J.* **204**, 649–667 (1976).

Field stimulated exoelectron emission from borosilicate glass

Terence M. Roberts & D. Brynn Hibbert

Department of Chemistry, Bedford College, University of London, Regent's Park, London NW1 4NS, UK

Exoelectrons, those electrons in energy levels of solids which are not fed by normal conduction mechanisms, have been detected at the surface of various materials[1]. Emission of exoelectrons may occur during mechanical stress[2], chemical reaction[3], or on heating[4] or irradiation by photons of suitable energy[5]. Thermally stimulated exoelectron emission gives rise to so-called glow curves, in which the emitted current goes through a maximum as the temperature is increased. Robertson[1] has suggested that an electric field may stimulate exoelectron emission in an analogous manner to thermal and photo emission. We describe here experiments which demonstrate the phenomenon of field stimulated exoelectron emission (FSEE), and discuss its relation to steady-state emission.

Electron emission from borosilicate glass under the influence of a strong electric field was first demonstrated by Hibbert and Robertson[6,7]. The emission current was continuous, noisy, of low intensity, and occurred from discrete 'active sites' on the glass surface. Further work[8] revealed that the total current is made up, in part, of fluctuating bursts from the active sites, whose characteristics are similar to each other. We have observed that on applying a high voltage to the glass emitter, a pulse of electrons is emitted, after which the steady-state emission is recorded. We suggest that the initial pulse of electrons is due to FSEE.

The apparatus (Fig. 1) for detecting FSEE was an evacuated (10^{-4} Pa) chamber containing a clean (trichloroethylene, chromic acid, distilled water), cylindrical Pyrex glass emitter surrounded by an earthed nickel gauze having 80 spaces per inch. The field at the surface of the glass for an applied voltage V is:

$$E = V\varepsilon_1/[r_1\varepsilon_2 \ln(r_1/r_0) + r_1\varepsilon_1 \ln(r_2/r_1)]$$

where r_0, r_1 and r_2 are the internal and external radii of the glass tube, and the radius of the nickel gauze respectively. ε_1 and ε_2 are the relative permittivities of glass and a vacuum ($\varepsilon_1 = 4.8$). For $r_0 = 5.5$ mm, $r_1 = 6.5$ mm and $r_2 = 9.5$ mm, $E = 370\,V$ V m^{-1}. The range of voltages applied in these experiments was 3–15 kV. Electrons which passed through the gauze were detected by a channel plate electron multiplier (Mullard G40/25) and a pattern of the emission from a section of the glass surface was displayed on a phosphor screen. A photomultiplier detected the light output from the screen. By suitable masking and focusing, the emission from a single active site was recorded. The photomultiplier output was calibrated and found to be proportional to the emission current in the range studied. Steady-state emission followed a Schottky-like equation[6]:

$$I = N \exp[-(\phi - aE^{1/2})/kT]$$

where N and a are constants (experimentally $a = 1.8 \times 10^{-4}$ eV V$^{-1/2}$ m$^{1/2}$), k is Boltzmann's constant, E the electric field strength, ϕ(= 0.97 eV) the energy difference between the electron levels and the vacuum level in the absence of an electric field, and T the absolute temperature. Exoelectron emission was studied by applying a voltage step to the glass emitter, and following the time development of the emission current on an oscilloscope. In addition, a ramped voltage was applied in an attempt to produce a glow curve of exoelectrons.

Figure 2 shows the steady-state emission pattern at 7 kV from a section of the borosilicate glass emitter. On stepping from 0 kV to the operating voltage each site showed a pulse of brightness which was recorded on the oscilloscope (Fig. 3).

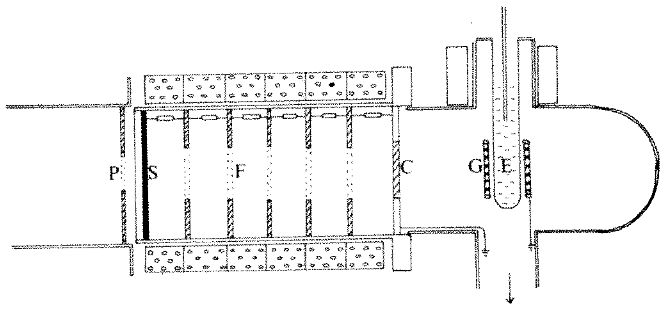


Fig. 1 Apparatus for detecting FSEE. E, emitter; G, nickel gauze; C, channel plate; F, focusing electrodes; S, phosphor screen; P, photomultiplier.

Fig. 2 Emission pattern of 13 mm diameter glass tube showing 2.5 mm diameter central area. Field at surface 2.59 MV m^{-1}.

0028-0836/82/180042—02$01.00

The height of the current pulse was dependent on the recovery time from switching off the previous voltage step. For a recovery time of <100 ms no exoelectron emission could be detected; ~10 s was required for the pulse to be restored to its maximum height (Fig. 4). The height of the pulse increased with both applied voltage and temperature. Such pulses were detected at the lowest voltage of 3 kV. A ramped voltage (0 V to 10 kV at 300 V s^{-1}) gave no glow curve. We attribute this to the ease at which exoelectrons are emitted by electric fields and the existence of a steady-state emission current which would mask any glow curve.

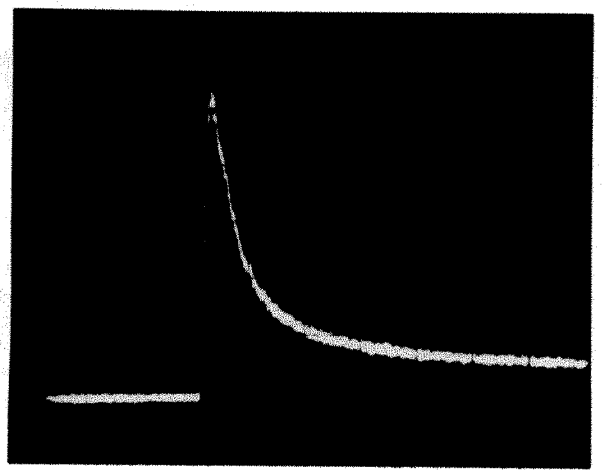


Fig. 3 Oscilloscope trace of the photomultiplier output of a burst of electrons from a single site, on stepping the emitter voltage from 0 V to 15 kV; 1 division on time axis = 0.5 s.

The connection between the exo and steady-state emission is still to be fully determined. Although steady-state emission may come from discrete energy levels in borosilicate glass[6] (there being no recognized conduction electrons), because of the slow, activated replenishment of exoelectrons, it is unlikely that the two types of emission occur from the same levels. Exoelectron emission has also been shown to depend strongly on surface conditions, and may disappear entirely in ultra-high vacuum[9]. Steady-state emission from glass is, however, maintained in ultra-high vacuum[7].

In conclusion, the pulse of electrons observed on applying a high voltage appears to be of exoelectrons; they are weakly held at the surface, and are only slowly replenished after emission. It should now be possible to investigate exoelectrons from other materials by the voltage step method. In particular, metals used as catalysts may be studied by this technique.

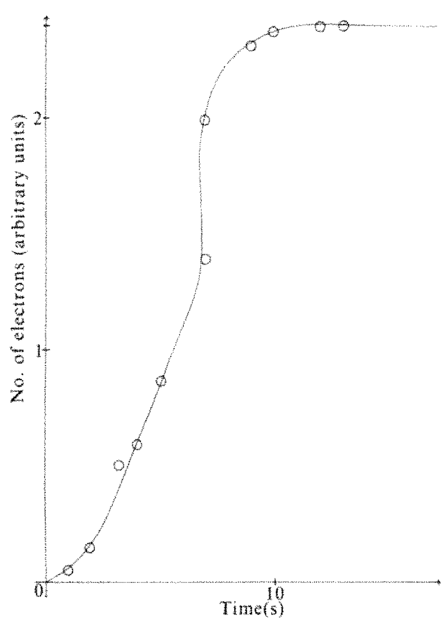


Fig. 4 Electron burst, measured as the height of the oscilloscope trace, as a function of the recovery time before a voltage step of 10 kV.

Received 8 February; accepted 12 March 1982.

1. Robertson, A. J. B. *Int. J. Electron* **51,** 607–619 (1981).
2. Balibar, F. *Recherche* **8,** 773–775 (1977).
3. Hoenig, S. A. & Utter, M. G. *J. Catal.* **47,** 210–213 (1977).
4. Shimida, H. & Nakajima, K. *Surface Sci.* **86,** 751–759 (1979).
5. Renfro, G. M. & Fishbeck, H. J. *Phys. Status Solidi* **A54,** 85–92 (1979).
6. Hibbert, D. B. & Robertson, A. J. B. *Proc. R. Soc.* **A349,** 63–79 (1976).
7. Hibbert, D. B. & Robertson, A. J. B. *Int. J. Electron* **48,** 301–303 (1980).
8. Hibbert, D. B. *27th int. Field Emission Symposium,* Tokyo, 138–139 (1980).
9. Kirihata, H. *Phys. Status Solidi* **A52,** K97–K99 (1979).

Helium isotopic systematics of oceanic islands and mantle heterogeneity

M. D. Kurz & W. J. Jenkins

Department of Chemistry, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

S. R. Hart

Center for Geoalchemy, Department of Earth and Planetary Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Isotopic variations in oceanic igneous rocks provide important constraints on models of oceanic mantle structure. Of particular interest is the global negative correlation between $^{87}Sr/^{86}Sr$ and $^{143}Nd/^{144}Nd$, which has been used to estimate 'bulk earth' values[1–3] for $^{87}Sr/^{86}Sr$, $^{87}Rb/^{88}Sr$ and $^{143}Nd/^{144}Nd$. Simple two-reservoir models have failed to explain all the isotopic variations, however, because of the complicated trends in Pb isotopes[4–6]. This has led to suggestions that recycled oceanic crust or sediments must be considered in these models[7–9]. We report here the results of helium isotopic analyses in basaltic phenocrysts from the islands of Gough and Tristan da Cunha. Because basalts from the islands lie near bulk earth on the Sr–Nd correlation diagram[3], the study was initiated to characterize the helium isotopic signature of this component. Whereas the ^{3}He in mantle gases is mostly primordial, the ^{4}He is primarily radiogenic, having been produced by decay of ^{238}U, ^{235}U and ^{232}Th. High $^{3}He/^{4}He$ ratios in igneous rocks therefore indicate primordial volatiles[10,11]. We believe that the present results are inconsistent with the notion that the mantle beneath Gough and Tristan da Cunha is primitive or undepleted relative to mid-ocean ridge basalt (MORB). Helium isotopic results on basaltic glasses and phenocrysts from the rift zone of Kilauea confirm the previously reported high values from this area[12–15]. We also report new analyses from Loihi Seamount (40 km south-east of Kilauea), which does seem to be derived from a more primitive source. When these data are combined with values for MORBs (from ref. 16) and plotted with respect to $^{87}Sr/^{86}Sr$, the observed trends offer insight into the different source regions for oceanic island basalts and the nature of mantle heterogeneity.

As sub-aerial basalts are largely degassed, they are of little use in the analysis of magmatic gas. This contrasts with MORB glasses, which are quenched rapidly enough on the ocean floor to trap substantial quantities of the magmatic gas[17,18]. Kaneoka *et al.*[12], however, have shown that the olivine phenocrysts from Kilauea basalts trap some of the magmatic gases that existed during crystal growth. The present results confirm the high $^{3}He/^{4}He$ ratios they reported for Kilauea phenocrysts, and we extend the approach to basalts from several other islands.

The samples were lightly crushed in a steel mortar, sieved, and the phenocrysts in the 20–40 mesh size range were separated by 'Frantz isodynamic' separation and hand-picking. Glassy samples were handpicked to exclude alteration, oxide

0028–0836/82/180043—05$01.00

crusts and non-vitreous chunks. Because the vesicles in submarine basalts can be a source of gas loss, chips >2 mm in size were selected[19]. After sonic cleaning, 1–2 g of the phenocrysts, or 400–500 mg of glass, were placed in a stainless-steel vessel designed specifically for *in vacuo* crushing[19]. The details of the gas mass spectrometry are similar to those reported earlier[19], except that the released helium was purified and expanded directly into the mass spectrometer, which resulted in a lower procedural blank ($1.0\text{–}2.0 \pm 0.3 \times 10^{-9}$ cm^3 STP ^{4}He with atmospheric $^3He/^4He$). Due to the small sample size, variability in the blank is the primary contributor to the experimental uncertainty (see Table 1).

Strontium isotopic analyses of several of the samples were performed using techniques that have been described elsewhere[20]. The measured $^{87}Sr/^{86}Sr$ ratios for Loihi Seamount samples KK 20-4 and KK 23-3 were both 0.70358 ± 0.00002, and Staudigel *et al.*[21] reported a value of 0.70358 ± 0.00004 for sample KK18-8. Gough Island samples ALR 26G and ALR 41G had $^{87}Sr/^{86}Sr$ ratios of 0.70527 ± 0.00004 and 0.70521 ± 0.00005 respectively; the Prince Edward sample WJ21E had a ratio of 0.70305 ± 0.00004. These values are reported relative to an Eimer and Amend standard value of 0.70800; the errors are 2σ for in-run statistics.

The helium results are reported in Table 1 for phenocryst samples from Tristan da Cunha, Gough, Prince Edward, Jan Mayen and Kilauea. We also report the analyses of submarine basaltic glass from the Loihi Seamount and the east rift of Kilauea, and the rock type and sample source for each sample are listed; a more detailed description is given in ref. 22.

Examination of the phenocryst samples in thin section shows that the most likely residence site for the helium is in the ubiquitous melt inclusions. In most cases these inclusions have undergone some post-entrapment crystallization, which explains why most of the helium is released by crushing *in vacuo* (see Table 1). Crystallization excludes the gas from the melt, but helium is still trapped in the phenocryst. The absence of xenocrysts in these samples was verified by petrographical examination[22].

In testing whether phenocrysts can be used for gas analysis, the samples from Kilauea were chosen because they are well characterized, and because they allow comparison of the phenocryst helium with the magmatic helium. The good agreement between the phenocrysts from the Kilauea picrite (H66050) and the two submarine glasses from the same volcanic rift supports the use of phenocrysts to indicate magmatic $^3He/^4He$ ratios. Our results also confirm the relatively high ratios in Kilauea phenocrysts reported by Kaneoka *et al.*[12], and by several laboratories for the Kilauea fumaroles[13,14]. The Loihi samples have even higher $^3He/^4He$ ratios, up to 31.9 times atmospheric, but have $^{87}Sr/^{86}Sr$ ratios that are indistinguishable from the values for the Kilauea samples[23]. Note that Kaneoka and Takaoka[15] have reported even higher $^3He/^4He$ ratios (up to 37 times atmospheric) for phenocrysts from Haleakala (Maui).

In contrast, the Tristan da Cunha and Gough samples contain helium with $^3He/^4He$ ratios significantly lower than the MORB values. As it was impossible to separate the amphibole in TK 26 from the interstitial opaque oxides and other accessory minerals, and because there is some question about the origin of these gabbroic nodules[24], we also analysed a basalt from Tristan da Cunha that contained large phenocrysts (TK 46A). Special care was taken to avoid opaque oxides and possible host rock contaminants (U and Th rich phases). The similar $^3He/^4He$ ratio for two different phenocryst phases from the same sample (TK 46A and ALR 26G) and between different volcanic eruptions suggests that this is not a problem.

While Kilauea, Loihi Seamount, Tristan da Cunha and Gough all have $^{87}Sr/^{86}Sr$ ratios higher (more radiogenic) than MORB values, the Hawaiian samples have higher $^3He/^4He$ ratios and Tristan and Gough have lower $^3He/^4He$ ratios. As shown in the plot of $^3He/^4He$ against $^{87}Sr/^{86}Sr$ (Fig. 1), the results fall into two distinct trends. For reference, we also show MORB

Table 1 Helium isotopic analyses for phenocryst and glass samples

Sample	Phase analysed	Rock type and location	^{4}He cm^3 STP g^{-1}	σ	^{3}He/^{4}He (R/R_a)	σ	Sample source
Tristan da Cunha							
TK 26	Amphibole	Gabbro xenolith in pyroclastics, Buff Gulch	3.3×10^{-7}	0.1	5.2	0.1	†
*TK 26	Amphibole		1.0×10^{-7}	0.1	4.7	0.1	†
TK 46A	Olivine	Ankaramite, near Big Point	7.3×10^{-9}	0.8	6.3	0.7	†
*TK 46A	Olivine		1.0×10^{-9}	0.1	5.6	2.1	†
TK 46A	Clinopyroxene		3.5×10^{-8}	0.2	5.1	0.3	†
Gough Island							
ALR 41G	Clinopyroxene	Highly pyroxene phyric basalt, Mount Powett	1.3×10^{-9}	0.4	4.9	1.6	‡
*ALR 41G	Clinopyroxene		$<7.0 \times 10^{-10}$	—	—	—	‡
ALR 41G	Clinopyroxene		5.8×10^{-10}	0.9	5.5	0.7	‡
ALR 26G	Olivine	Ankaramite, Mount Powett	1.53×10^{-8}	0.06	6.2	0.2	‡
ALR 26G	Clinopyroxene		3.93×10^{-8}	0.09	6.2	0.3	‡
Jan Mayen							
JM 151A	Olivine	Ankaramite	—	—	6.3	0.5	§
	Clinopyroxene		9.7×10^{-9}	0.4	6.8	0.3	
Prince Edward							
WJ 21E	Olivine	Ankaramite, top of western escarpment	2.33×10^{-8}	0.05	7.4	0.2	‖
Kilauea							
H66050	Olivine	Picrite, crater wall	5.8×10^{-9}	0.3	14.0	1.4	¶
Puna 2	Glass	Tholeiite, East Rift	1.51×10^{-7}	0.03	14.7	0.5	#
Puna 8	Glass	Tholeiite, East Rift	1.88×10^{-7}	0.04	14.5	0.3	#
Loihi Seamount							
KK 23-3	Glass	Tholeiite	6.1×10^{-8}	0.04	23.1	0.8	††
KK 20-4	Glass	Alkali basalt	5.20×10^{-7}	0.14	24.1	0.7	††
KK 18-8	Glass	Tholeiite	2.71×10^{-7}	0.06	31.9	0.7	††

All samples were crushed *in vacuo* except those denoted*, which had the helium extracted by melting *in vacuo* after crushing. All $^3He/^4He$ ratios are reported relative to atmospheric (R/R_a) using an atmospheric value of 1.384×10^{-6}. Sample sources: †Dr S. Humphris; ‡Dr A. le Roex; §Dr S. Maaloe; ‖Dr W. J. Voerwoerd, see ref. 50; ¶Dr S. O. Agrell; #Dr J. Moore, see refs 51 and 23, sample numbers refer to Table 2 of ref. 51; ††Dr D. Clague, see ref. 52.

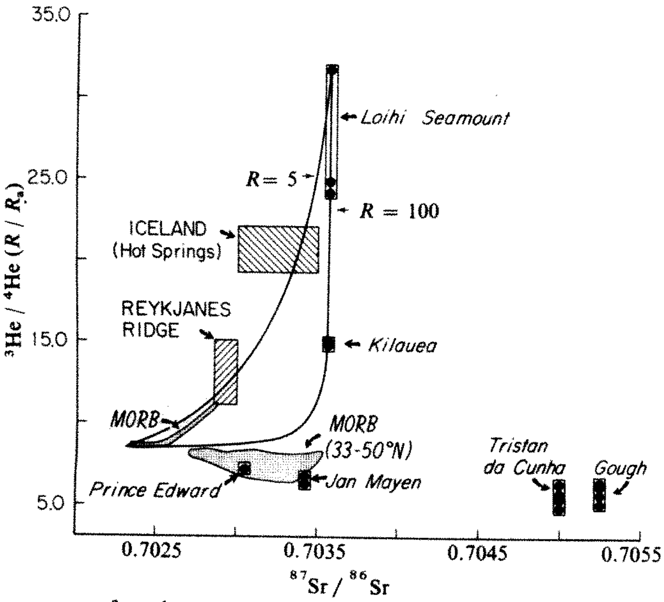


Fig. 1 $^{3}He/^{4}He$ (relative to atmosphere) plotted against $^{87}Sr/^{86}Sr$ for oceanic volcanic rocks. Data sources: Mid-Atlantic Ridge (MORB)[16]; Reykjanes Ridge[26,40]; Iceland[26,27,40,41]; Kilauea this study and ref. 23; Loihi Seamount, this study; Prince Edward, this study; Jan Mayen this study and ref. 42; Tristan da Cunha this study and ref. 3; Gough Island, this study. Note that the Gough, Prince Edward, MORB, Kilauea, and Loihi samples had helium and strontium isotopic analyses run on the same samples; in all other cases, the fields indicated represent ranges of values for similar samples from the references listed above. For the Icelandic hot springs, the highest reported $^{3}He/^{4}He$ ratios[26,27] were selected to minimize atmospheric effects. The highest terrestrial $^{3}He/^{4}He$ ratios (37 × atmospheric) reported by Kaneoka and Takaoka[15] are not plotted because they do not report $^{87}Sr/^{86}Sr$ ratios. The two mixing lines were calculated assuming component 1 has $^{3}He/^{4}He = 8.5 \times R_a$, $^{87}Sr/^{86}Sr = 0.70230$ and component 2 has $^{3}He/^{4}He = 32.0 \times R_a$, $^{87}Sr/^{86}Sr = 0.70358$. As shown by Langmuir *et al.*[28], curvature is determined by the ratio R, where in the present case:

$$R = \frac{{}^{4}He_1 \, {}^{86}Sr_2}{{}^{4}He_2 \, {}^{86}Sr_1}$$

with age of the Earth. This contrasts with continental crust, which is greatly enriched in U and Th and degassed of ^{3}He values for the North Atlantic, which are discussed in detail elsewhere[16], and literature values for Icelandic hot springs and the Reykjanes Ridge[25,26].

To explain these trends, we can immediately eliminate *in situ* (post-eruptive) decay of U and Th isotopes to produce the low $^{3}He/^{4}He$ trend, as all the samples are of essentially zero age and the phenocrysts analysed contain low U contents. This is also supported by the analyses of olivine and clinopyroxene from the same sample, which yielded similar results. We believe that the trends in Fig. 1 are most easily accounted for by three component mixing. One component, with $^{3}He/^{4}He$ of $\sim 1.17 \times 10^{-5}$ (8.5 × atmospheric) and $^{87}Sr/^{86}Sr$ of ~0.7025, would then be identified as the 'typical MORB' reservoir. A second component, characterized by high $^{3}He/^{4}He$ and higher $^{87}Sr/^{86}Sr$ ratios, would be consistent with a more primitive source derived from 'mantle plumes', as has been suggested for Hawaii and Iceland[15,27]. The third component is characterized by low $^{3}He/^{4}He$ and high $^{87}Sr/^{86}Sr$ ratios, as indicated by the results for Tristan and Gough.

The curve defined by the MORB–Loihi trend in Fig. 1 is consistent with mixing, as ratio–ratio plots do not always display straight lines for binary mixing[28]. Several mixing lines (calculated for different He and Sr concentrations in the end-members) are shown for reference in Fig. 1. The helium isotopic ratio of a possible primordial end-member (either 'planetary' or solar) is shown in Fig. 2. Because of the high He/(Th + U) ratio characteristic of the Sun and meteorites, there is little change in the $^{3}He/^{4}He$ ratio (due to addition of radiogenic ^{4}He) during formation, resulting in quite low present-day $^{3}He/^{4}He$ ratios (see Fig. 2). The primordial end-member would then lie between the Loihi seamount point and bulk earth as plotted in Fig. 2. Because it is not clear that there is a genetic relationship between the Earth and the meteorites, or whether any undifferentiated terrestrial mantle still exists, the true end-member must remain uncharacterized[16].

The low $^{3}He/^{4}He$ samples for Gough and Tristan da Cunha require mixing with a reservoir that has been enriched in Th and U relative to ^{3}He for time periods long enough to lower the $^{3}He/^{4}He$ ratio. The time required depends on the $^{3}He/(Th + U)$ ratio[16]. Our data cannot distinguish between seawater, subducted oceanic crust plus sediment, or old continental crust as a source for this component. As shown in Fig. 2, any of these would serve as an appropriate end-member if the mixing hypothesis is used to explain the variations. However, given the Sr-Nd correlation and the expected effect of seawater addition on these isotopes, seawater seems unlikely[29]. Seawater also contains small quantities of helium relative to basaltic melts, so addition of large quantities would be required to lower the $^{3}He/^{4}He$ ratio, and would result in extreme variations in $^{87}Sr/^{86}Sr$.

The suggestion[30] that an important source of isotopic variation in oceanic basalts is contamination from the oceanic crust through which the eruptive basalts must pass is plausible in that oceanic crust should separate ^{3}He from Th and U by degassing. However, this mechanism seems unlikely for several reasons. First, volcanics on Tristan da Cunha, Gough, Jan Mayen and Prince Edward are erupted through oceanic crust that is much younger (and thinner) that the crust beneath Hawaii, and yet the $^{3}He/^{4}He$ ratios are lowest. If contamination were significant, one would expect the oldest, most radiogenic crust to lower the ratio the most. Second, special conditions would be required to lower the magmatic $^{3}He/^{4}He$ ratio by this mechanism. The contaminating crust must have lost most of its initial He, and must have produced (and retained) substantial quantities of radiogenic ^{4}He. The generally low U and Th contents of unaltered oceanic crust require extreme enrichment in these

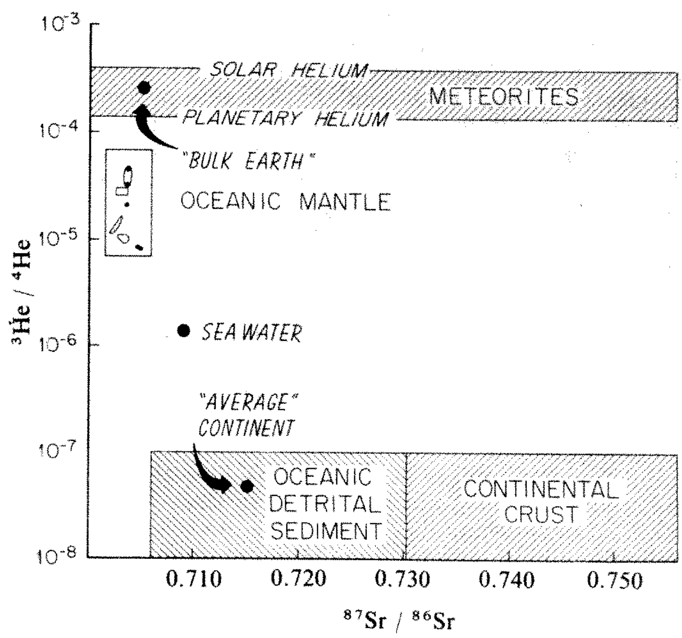


Fig. 2 Simplified plot of $^{3}He/^{4}He$ versus $^{87}Sr/^{86}Sr$ for terrestrial materials and meteorites. Data sources: chondrites[43,44]; bulk earth[3,44]; oceanic mantle (see Fig. 1); seawater[45,46]; continental shield[10,47]; oceanic detrital sediments[48]; and average crust[6,49]. Oceanic detrital sediments, and average crust are assumed to have the same range of $^{3}He/^{4}He$ ratios as continental gases[10]. We have also ignored meteoritic spallation helium, which can have $^{3}He/^{4}He$ ratios higher than 10^{-4}. The upper boundary for the oceanic mantle field is defined by helium analyses by Kaneoka and Takaoka[15].

elements to generate enough ^4He in reasonable time periods. This is particularly true for Tristan de Cunha and Gough, which are situated on oceanic crust that is 10–20 Myr old.

One mechanism that cannot be ruled out is the separation of He from U and Th by multiple melting events. For example, if the mantle beneath Tristan da Cunha and Gough were 'enriched' by the addition of a small amount of melt or fluid (equivalent to metasomatism[31]), it is conceivable that helium would be lost by degassing, while the U and Th would be retained. The $^3He/^4He$ ratio would then decrease, due to extremely low $^3He/U$ and $^3He/Th$ ratios; on melting, this enriched mantle would yield low $^3He/^4He$ ratios. Trace element analyses from Tristan da Cunha and Gough suggest that the lavas are enriched in Th and U relative to chondrites[32], making this feasible. However, the small degree of partial melting, which may generate oceanic island alkali basalts, makes it difficult to determine the source mantle characteristics from the trace element concentrations and ratios[33]. In addition, the physical process by which helium is lost from the mantle without melt removal is unclear.

The subduction of altered oceanic crust and sediments into the mantle has been suggested as an explanation for some of the Sr–Nd–Pb variations[7,8,34], and is quite consistent with our results for Tristan da Cunha and Gough. If the low $^3He/^4He$, high $^{87}Sr/^{86}Sr$ mantle source region is produced by adding subducted crust into the mantle and then remelting, it is important to evaluate the effect on helium isotopes. The $^3He/^4He$ ratio that would result from remelting of this material is a function of the initial helium content of the crust, the amount of degassing it has undergone (both before and during subduction), the helium content of the mantle to which it is added, and the relative proportions of the two mantle types. These formidable uncertainties make a quantitative treatment impossible; however, there is clearly sufficient U and Th present to produce the observed variations, depending on the physical processes occuring. For example, in 200 Myr, $4\times10^{-6}\,cm^3g^{-1}$ of radiogenic ^4He will accumulate in oceanic crust having an average U content of 100p.p.b. (parts per 10^9) and $Th/U=2$ (ref. 35). If we take $1\times10^{-5}\,cm^3g^{-1}$ as an upper limit of the initial helium content of oceanic crust, the $^3He/^4He$ ratio in this crust will decrease by at least 30% in 200 Myr. In addition, it is possible that hydrothermal alteration adds U to the crust, making 100 p.p.b. a lower limit[7,36,37]. The other extreme is an oceanic crust that degasses completely on formation, leaving only radiogenic ^4He to remix with the mantle on resubduction. Craig and co-workers[38] have interpreted low $^3He/^4He$ ratios in present-day back-arc volcanic systems to be a result of mixing between the helium in the downgoing slab and the helium in the underlying mantle, which suggests that degassing continues after crustal formation. In both cases, the $^3He/^4He$ will decrease to some extent, depending on the mixing ratio of the two components. Clearly, there is sufficient U and Th to lower the $^3He/^4He$ ratio significantly even before subduction. If the subducted crust remains in the mantle for long periods, as suggested by Hofmann and White[7], substantial quantities of radiogenic ^4He may accumulate.

All three mechanisms described above (crustal contamination, multiple melting, and remelting of subducted crust in the mantle) could conceivably produce the low $^3He/^4He$ trend defined by the Tristan da Cunha and Gough points. At present, we believe that the subducted crust hypothesis most easily accounts for the trend. In contrast to the other two processes, subduction is a commonly observed phenomenon. The mechanism for separation of He from U and Th is degassing of the oceanic crust, another presently observable phenomenon[14]. Further, the allowed time periods and U enrichments necessary to lower the $^3He/^4He$ ratio are geologically quite reasonable.

If all the high $^3He/^4He$ islands are produced by two component mixing with the same end-members, then various $^4He/^{86}Sr$ ratios are required, as this will determine the curvature of the mixing line. For example, if the Kilauea isotopic signature is derived by mixing the two end-members shown in Fig. 1, then the high $^3He/^4He$ end-member must have higher Sr contents or lower He contents (or both, see Fig. 1). Alternatively, another end-member may be involved. Note that Tatsumoto[35] observed different Pb isotopic compositions for each of the five sub-aerial Hawaiian volcanoes. He suggested that the linear trend, defined by the Hawaiian volcanoes, on the $^{207}Pb/^{204}Pb$ against $^{206}Pb/^{204}Pb$ plot was a mixing line. To test the mixing hypothesis, a more detailed study of the Loihi Seamount and the island of Hawaii is underway in our laboratories.

It would appear that $^3He/^4He$ measurements, coupled with the other isotopic measurements, are an important discriminant between primordial and 'recycled' mantle source regions. The reason for the large variations is that separation of He and Th + U occurs by degassing, which is not the case for the Rb–Sr, Nd–Sm, and U–Pb systems. Note that the samples lying close to bulk earth on the Nd–Sr correlation line, such as Tristan da Cunha and Gough, are not necessarily representative of undepleted mantle. It is possible to derive the trend by mixing between some crustal components[9,39] and the depleted MORB reservoir, resulting in a coincidental bulk earth value.

Therefore, on the basis of the helium isotopic information, three distinct mantle reservoirs are required: depleted (the MORB source), undepleted, and recycled. The islands displaying the highest $^3He/^4He$ ratios have tholeiitic affinities (Hawaii and Iceland), while the islands consisting of alkali basalts have characteristically lower $^3He/^4He$ ratios (Tristan da Cunha and Gough). More detailed sampling is required to confirm this trend.

We thank the following for providing samples: S. O. Agrell, D. Clague, S. Humphris, A. le Roex, S. Maaloe, J. Moore, R. K. O'Nions and W. Voerwoerd; also A. le Roex, G. Thompson, J.-G. Schilling and H. Dick for interesting discussions, and B. Haskell for editorial comments. M.K. thanks the Education Office of the WHOI/MIT Joint Program in Oceanography for their continuing support. This work was supported by NSF grant OCE-7909379 and USGS contract 14-08-001-G541. This is WHOI contribution no. 4955.

Received 24 August 1981; accepted 5 March 1982.

1. Richard, P., Shimizu, N. & Allègre, C. J. *Earth planet Sci. Lett.* **31**, 269 (1976).
2. DePaolo, D. J. & Wasserburg, G. J. *Geophys. Res Lett.* **3**, 249 (1976).
3. O'Nions, R. K., Hamilton, P. J. & Evensen, N. M. *Earth planet. Sci. Lett.* **34**, 13 (1977).
4. Sun, S. S. *Phil. Trans. R. Soc.* **A297**, 409–447 (1980).
5. Allègre, C. J., Brévart, O., Dupre, B. & Minster, J. F. *Phil. Trans. R. Soc.* **A297**, 447 (1980).
6. O'Nions, R. K., Evensen, N. M. & Hamilton, P. J. *J. geophys. Res.* **84**, 6091 (1979).
7. Hofmann, A. W. & White, W. M. *Yb. Carnegie Instn Wash.* 477–483 (1980).
8. Chase, C. G. *Earth planet. Sci. Lett.* **52**, 277–284 (1981).
9. Anderson, D. L. *Science* **213**, 82–89 (1981).
10. Tolstikhin, I. N. in *Terrestrial Rare Gases* (ed. Alexander, E. C.) (Center for Academic Publications, Japan, 1978).
11. Craig, H. & Lupton, J. in *The Sea* Vol. 7, 391–428 (Wiley, New York, 1981).
12. Kaneoka, I. & Takaoka, N. *Earth planet. Sci. Lett.* **39**, 382 (1978).
13. Craig, H. & Lupton, J. *Earth planet. Sci. Lett.* **31**, 369–385 (1976).
14. Jenkins, W. J., Edmond, J. M. & Corliss, J. B. *Nature* **272**, 156–158 (1978).
15. Kaneoka, I. & Takaoka, N. *Science* **288**, 1366–1368 (1980).
16. Kurz, M. D., Jenkins, W. J., Schilling, J.-G. & Hart, S. R. *Earth planet. Sci. Lett.* **58**, 1 (1982).
17. Funkhouser, J. G., Fisher, D. E. & Bonatti, E. *Earth planet. Sci. Lett.* **5**, 95–100 (1968).
18. Dymond, J. & Hogan, L. *Earth planet. Sci. Lett.* **20**, 131 (1973).
19. Kurz, M. D. & Jenkins, W. J. *Earth planet. Sci. Lett.* **53**, 41–54 (1981).
20. Hart, S. R. & Brooks, C. *Geochim. cosmochim. Acta* **38**, 1799–1806 (1974).
21. Staudigel, H. *et al. EOS* **62**, 1087 (1981).
22. Kurz, M. D. thesis, Woods Hole Oceanographic Inst./MIT Joint Program (1982).
23. Hart, S. R. *Earth planet. Sci. Lett.* **20**, 201–203 (1973).
24. Le Maitre, R. W. *Miner. Mag.* **37**, 185 (1969).
25. Polak, B. G., Komonmov, V. I., Tolstikhin, I. N., Mamyrin, B. A. & Khabarin, L. *Publ.* 119, *Int. Ass. Hydrol. Sci.*, 117 (1975).
26. Poreda, R., Craig, H. & Schilling, J.-G. *EOS* **61**, 1158 (1980).
27. Craig, H., Lupton, J., Welhan, J. & Poreda, R. *Geophys. Res. Lett.* **5**, 897 (1978).
28. Langmuir, C. H., Vocke, R. D., Hanson, G. N. & Hart, S. R. *Earth planet. Sci. Lett.* **37**, 380–392 (1978).
29. O'Nions, R. K., Carter, S. R., Cohen, R. S., Evenson, N. M. & Hamilton, P. J. *Nature* **273**, 435–438 (1978).
30. O'Hara, M. J. *Phil. Trans. R. Soc.* **A297**, 215–227 (1980).
31. Lloyd, F. E. & Bailey, D. K. *Phys. chem. Earth* **9**, 389–446 (1975).
32. Kable, E. J. thesis, Univ. Capetown (1974).
33. Gast, P. *Geochim. cosmochim. Acta* **32**, 1057 (1968).
34. Zindler, A. *Chapman Conf. on Oceanic Crust*, April (1981).
35. Tatsumoto, M. *Earth planet. Sci. Lett.* **38**, 63–87 (1978).
36. MacDougall, J. D. *Earth planet. Sci. Lett.* **35**, 65–70 (1977).
37. Mitchell, W. S. & Aumento, F. *Init. Rep. DSDP* Log **37**, 547–559 (1974).
38. Craig, H., Lupton, J. & Weiss, R. in *Terrestrial Rare Gases* (ed Alexander, E.) (Japan Scientific Society, Tokyo, 1978).
39. Carter, S. R., Evensen, H. M., Hamilton, P. J. & O'Nions, R. K. *Science* **202**, 743–747 (1978).
40. Hart, S. R., Schilling, J. G. & Powell, J. L. *Nature* **246**, 104 (1973).
41. Zindler, A., Hart, S., Frey, F. & Jakobbson, S. *Earth planet. Sci. Lett.* **45**, 249–262 (1980).

42. O'Nions, R. K. & Pankhurst, R. J. *J. Petrol.* **15**, 603 (1974).
43. Mittlefehldt, D. W. & Wetherill, G. W. *Geochim cosmochim. Acta* **43**, 201 (1979).
44. Heymann, D., in *Handbook of Elemental Abundances in Meteorites* (ed. Mason, B.) (Gordon & Breach, Edinburgh, 1971).
45. Clarke, W. B., Jenkins, W. J. & Top, Z. *Int. J. app. Rad. Isotopes* **27**, 515 (1976).
46. Faure, G. & Powell, J. L. *Strontium Isotope Geology* (Springer, Berlin, 1972).
47. McColloch, M. T. & Wasserburg, G. J. *Science* **200**, 1003–1011 (1978).
48. Dasch, E. J. *Geochim. cosmochim. Acta* **33**, 1521–1552 (1969).
49. Hart, S. R. & Allegre, C. J. in *Physics of Magnetic Processes* (ed. Hargreaves, R.) 121–151 (Princeton University Press, 1980).
50. Voerwoerd, W. J. in *Marion and Prince Edward Islands* (eds Van Zinderen, E. M. *et al.*) (Balkema, Capetown, 1971).
51. Moore, J. G. *Am. J. Sci.* **263**, 40–52 (1965).
52. Moore, J., Claque, D. & Normark, W. *Geology* **10**, 88–92 (1982).

Scaling rules in rock fracture and possible implications for earthquake prediction

C. J. Allègre, J. L. Le Mouel & A. Provost

Institut de Physique du Globe, Université Paris 6, 2 place Jussieu, Paris 75005, France

A major preoccupation in physical sciences has been to interpret macroscopic events from microscopic phenomena. In some cases the change of scale is efficient and fairly easy to perform, but in others it turns out to be difficult and uninteresting. Success or failure is due more to the nature of the events than to the efficiency of the theoretical methods used. Some macroscopic phenomena have their origin in a microscopic organization which can be transferred to larger scales whereas others attain their structure on the macroscopic scale itself. Thus before applying scaling laws techniques[1–4] one must ensure that embedded scales are suggested by physical observations. That this seems to be the case for the fracture of rocks is supported by geological, seismological and rock mechanics observations. We have therefore built a very simple model based on scaling laws which yields a criterion for fragility at different scales and views rupture as a critical point. We use this model here to outline a general approach to earthquake prediction.

Fracturing occurs in rocks at all scales, from the microscale (microcracks) to the continental scale (megafaults), and the geologist can equally well observe embrittlement and rupture phenomena under the microscope as on satellite photographs (Fig. 1). But are the various scales of fracture related to one another? The following observations suggest they are: (1) field geologists know that the great faults of the crust—such as the San Andreas fault—actually consist of anastomosed faults, sometimes arranged *en échelon*, thus weakening a whole domain of the crust, down to variable depths[5]. (2) Seismologists who study source phenomena often have to introduce in their models complex rather than single faults, each one contributing to the observed radiation pattern[6,7]. (3) Rock mechanicists, when studying fracture in the laboratory, observe that it is preceded by the concentration of a swarm of microfissures which are themselves the result of an accumulation of microcracks[8–11]. From this set of observations one can suggest that fracture at the macroscopic scale is a consequence of accumulations of ruptures at lesser scales.

This hypothesis has been actually adopted by Brace and his students[8,9,11,12]. They have submitted rock samples to progressive triaxial loading ($\sigma_1, \sigma_2, \sigma_3$) and studied the increase of microcrack density with increasing load. This increase seems to be the result of two distinct processes: (1) the nucleation of new cracks, that is, the birth of new rupture points in the material; and (2) the growth of pre-existing cracks. In fact, as noted by Brace *et al.*[8] and Tapponnier and Brace[11], the nucleation seems to occur most often from a pre-existing crack. The distinction between the two processes is thus subtle and the increase in crack density can be considered to be ruled by a single phenomenon with a given activation energy.

The law governing the increase of microcrack density with deviatoric stress ($\sigma_1 - \sigma_3$) depends strongly on the confining pressure. But what seems to occur generally is that the macroscopic fracture is not preceded by an accelerated growth of microcrack density as measured over the whole sample. When the sample is examined at different scales one observes that cracks collect in some regions, but that these microscopic regions are roughly homogeneously distributed in the medium, even when the rupture threshold has been reached. On the other hand, the larger the scale the stronger is the spatial heterogeneity. The heterogeneity reaches, of course, a limit which determines the fracture itself whose orientation follows the laws established by Anderson[13].

We will now try to explain those observations with a simple renormalization group (RG) model and examine possible implications of the model for earthquake prediction.

For each elementary domain of rock (say 100 μm), we define two states: when the local microcrack density in the domain is greater than some critical value, it is considered as fragile (f); otherwise it is considered as sound (s).

As shown elsewhere (for example, see ref. 11), the mean microcrack density d depends linearly on ($\sigma_1 - \sigma_3$) (for a given σ_2), but the local density varies considerably within the sample. The probability for an elementary domain to be fragile, p, is directly proportional to d and thus linearly related to ($\sigma_1 - \sigma_3$) (for a given σ_2):

$$d = d_0 + b(\sigma_1 - \sigma_3)$$

$$p = ad = ad_0 + ab(\sigma_1 - \sigma_3) \qquad (1)$$

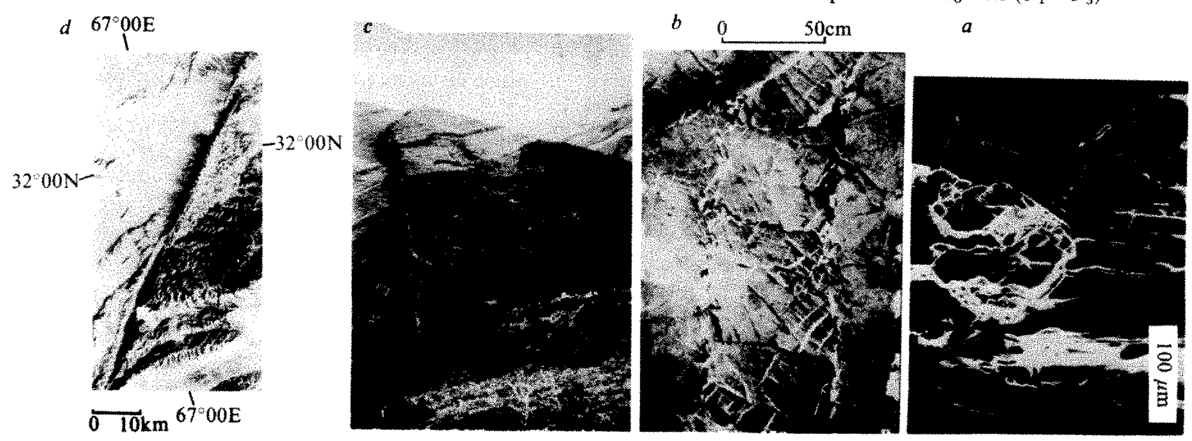


Fig. 1 Examples of fracture at different scales. *a*, Microcracks in quartz grains are induced by intense cracking in magnetite and plastic flow in biotite. (Westerly granite, fracture stress, 35 bar of confining pressure, room temperature.) *b*, Tension gashes, stylolites and micro shear faults in horizontal Mesozoic limestones (near Les Matelles, in Languedoc, France). The microstructures combine to form a fault zone at a larger scale. *c*, Master fault of the El Asnam, Algeria, earthquake (magnitude = 7.3, 10 October 1980). In the hills north of El Attaf, the break, several kilometres long, has up to 4 m of vertical throw. *d*, Landsat image of the Chaman strike-slip fault south-west of the Katawaz basin near the border between Afghanistan and Pakistan. The fault system, ~1,200 km long, may have accommodated as much as 500 km of left lateral displacement of India past Afghanistan in the past 40 Myr.

0028–0836/82/180047—03$01.00

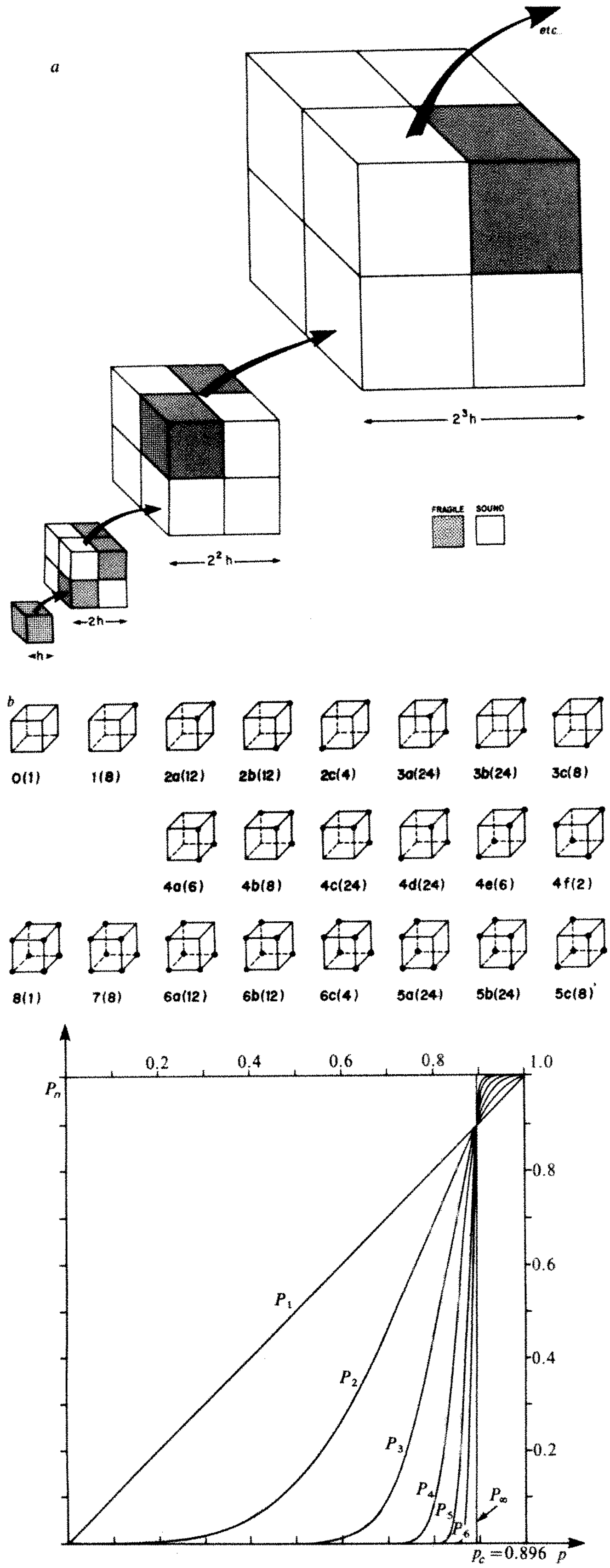


Fig. 2 *a*, Illustration of the renormalization group model. *b*, Now an *n*-order cube is depicted as a simple cube with its corners representing the eight constitutive $(n-1)$-order cubes. Each corner is marked by a solid dot when the corresponding $(n-1)$-order cube is fragile. This defines 22 configurations labelled 0, 1, 2a, . . . 8 with their multiplicities in parentheses. The probability for a configuration with m corners fragile and multiplicity μ is: $\mu(P_{n-1})^m(1-P_{n-1})^{8-m}$. *c*, Probability P_n for an n cube being fragile plotted against the microscopic probability p.

Parameters d_0 and b depend on σ_2, whereas a can be considered purely geometrical.

Let us now develop a criterion for fragility at different scales. Let the elementary domain of order 1 be a cube of unit side (for example, 100 μm). Let the order of 2 cube be composed of 2^3 elementary order of 1 cubes, and so on (Fig. 2*a*).

In terms of $(n-1)$ cubes (we write n cube for the order of n cube), there are $2^8 = 256$ different n cubes. Taking the symmetry group of the cube into consideration, this number reduces to 22 topologically different configurations, each with a given multiplicity (Fig. 2*b*). An n cube will be considered sound whenever there is a 'pillar' of sound $(n-1)$ cubes that links two opposite faces. We thus consider as fragile the configurations labelled 4f, 5c, 6b, 6c, 7 and 8 in Fig. 2*b*. If P_n is the probability for an n cube being fragile, then:

$$P_1 = p, \quad P_2 = \Pi(p), \quad \ldots, \quad P_n = \Pi(P_{n-1}) \qquad (2)$$

with

$$\Pi(x) = 2x^4(1-x)^4 + 8x^5(1-x)^3 + 16x^6(1-x)^2 + 8x^7(1-x) + x^8$$

The polynomial $\Pi(x)$ has the following simple properties: $\Pi(x) = x$ for $x = 0$, $x = p_c = 0.896$ and $x = 1$. It monotonously increases with x on $[0, 1]$, being less than x on $]0, p_c[$ and greater than x on $]p_c, 1[$. Furthermore, $\Pi'(1) = \Pi'(0) = \Pi''(0) = \Pi'''(0) = 0$. This explains the characteristics of the probability law P_n, which are illustrated in Fig. 2*c*. A striking feature is how rapidly the curves converge to the (Heaviside) step-function $H(p-p_c)$ with increasing n. For $n = 5$, P_n positively jumps from 0 to 1 when p passes p_c.

This crude model allows one to account very simply for the observations of Tapponnier and Brace[11]. When counting microcrack density at the lowest scale an increase of this density increases with $(\sigma_1 - \sigma_3)$ is observed, but nothing clearly announces the macroscopic rupture which appears as an 'exogen' discontinuity in a continuous process. The situation is quite different when looking at larger scales: little occurs before p gets to values close to p_c; then the fragility probability P_n jumps rapidly to values close to 1 and rupture occurs (see Fig. 3). p_c is thus a critical probability, the threshold for crack percolation.

Of course, the fragility criterion we have chosen is arbitrary and oversimplified. Other configurations of $(n-1)$ cubes may be supposed to make the n cube fragile. More refined criteria may be built by considering higher-order scaling steps ($3^3 = 27$, $4^3 = 64$, and so on, elementary cubes instead of $2^3 = 8$). In any case we get a recursion relationship between P_{n-1} and P_n of the form $P_n = \Pi(P_{n-1})$, Π being a polynomial with properties analogous to the ones above. Figure 4 illustrates a case in which the renormalization polynomial $\Pi(x)$ has no fixed point except the trivial ones $x = 0$ and $x = 1$. Although such a choice might seem inappropriate ($P_n \to 0$ when $n \to \infty$ for all $p < 1$), it leads to the same conclusions (a step from very low to very high probabilities at the highest scales) due to the fact that there is

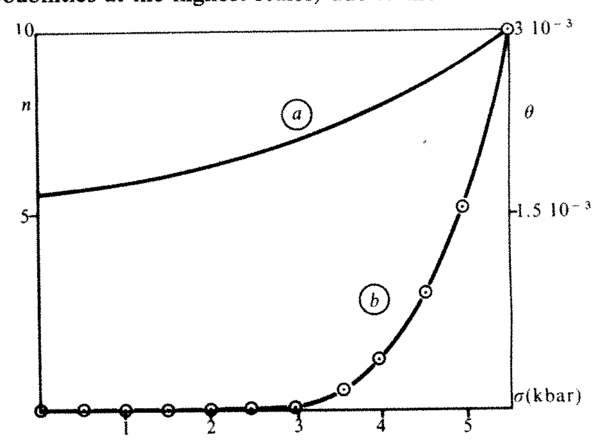


Fig. 3 Schematic representation of some experimental results (adapted from ref. 11). Sample of Westerly granite, confining pressure 500 bar. *a*, Variation with σ of a microscopic property (number of crack intersections/mm, n); *b*, variation with σ of a macroscopic property (volumetric crack dilatancy, θ).

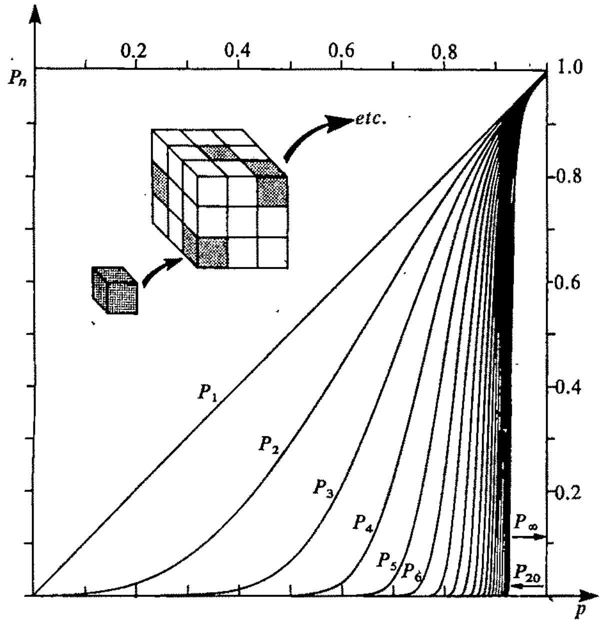


Fig. 4 Now an n cube, made of 27 $(n-1)$ cubes, is considered as fragile when three fragile $(n-1)$ cubes are aligned along one of the three medians of the n cube. Then $P_1=p$, $P_2=\Pi(p),\ldots, P_n=\Pi(P_{n-1})$ with $\Pi(x)=3x^3-3x^5+x^7$. With increasing n, the P_n curves sharpen rapidly to a quasi-step, then shift only slowly to the right. The last curve drawn is that for $P_{20}(3^{19}>10^9)$. This shows that the prediction for megafaulting is not significantly altered by the mathematical absence of a critical probability p_c.

not an infinity of scaling steps from microcracks to megafaults ($100\ \text{km}/100\ \mu\text{m}=10^9$). On the same basis it is also possible to build 'anisotropic models' in which the fragility is relative to a given direction and varies with this direction. But rather than generalizing, we will now draw some inferences about earthquake prediction from the simple model considered above.

The dilatancy theory for earthquake prediction which has been developed from laboratory experiments[8,14,15] has had some success but has failed to provide a general solution. Using a similar approach, we now show how to draw, from laboratory experiments, other useful inferences on earthquake prediction.

Earthquakes are fracture phenomena which most often occur on pre-existing sets of faults. Together with many theoretical seismologists we may assume that the reactivation of a fault system is a fracture phenomenon polarized by the existence of the fault itself[16,17]. Then the approach we propose, which determines a critical fragility, can also apply to earthquakes. One only needs to consider that the macroscopic domain with which our computations are concerned is the domain of the faults.

Earthquakes are probably generated by a strain accumulation under a constant deviatoric stress $\sigma=\sigma_1-\sigma_3$ rather than by a continuous increase of σ with time. Mogi[18] has proposed that, in a medium submitted to a constant σ, the increase in the probability of occurrence of fracture during the time interval $(t, t+dt)$ does not depend on t and is given by $dp=\mu_0 \exp(\beta\sigma)\, dt$, where μ_0 and β are constants. We can also consider the more general case when stress is increasing with time:

$$p(t)=p_0+\int_0^t \mu_0 \exp(\beta\sigma(t))\,\mathrm{d}t \qquad (3)$$

(We may assume, in such conditions, that the microcrack density obeys a zero-order kinetic equation $\mathrm{d}'(t)=\delta_0 \exp(\beta\sigma)$, $\mathrm{d}(t)$ and $p(t)$ being related by $p(t)=a\,\mathrm{d}(t)$, as above.)

It is then possible to compute the curves $P_n(t)=P_n[\sigma(t)]$ by the recursion formula (2). Of course, the shape of those curves is the same as that in Fig. 4, the parameter along the horizontal axis now being the time. For example, in the experiments of Kranz on samples of Barre granite[19], fracture occurs about 150 s after a stress of 2 kbar has been applied. To get the time parametrization, one has only to replace the $(0, p_c)p$ interval by a (0, 150 s) t interval. If σ is supposed to vary with time, a variable contraction or dilation of the horizontal axis can modify the figure of the $P_n(t)$ curves.

Suppose now we have two physical parameters, (x) and (y), which depend in a known way on crack density at different scales n_1 and n_2. It is then possible to predict the critical point $p(t)=p_c$ from the predicted intersection of the two curves $x(t)$ and $y(t)$. So it may be thought that the prediction of earthquakes demands the observation of phenomena characterized by different scales. For example, if one studies seismic waves for periods of 0.1, 1, 10, 100, ... s, one may think that the values of the corresponding attenuation ratios will tend towards some limit at the critical point (the partial failure of the (V_P/V_S) ratio prediction technique may come from the fact that in some circumstances this ratio depends on the scale but not in others, whether pores containing water in the rock are connected or not).

The crude model we propose here may be tested in two ways. First by studying in the laboratory the variation of crack density at different scales as $(\sigma_1-\sigma_3)$ increases; second, by studying the variations of some physical properties which depend on crack density at different scales, especially in the vicinity of fracture.

Despite the extreme simplicity and the limits of our approach, if our basic conjecture is correct, a set of prediction methods could be built up.

We thank T. Madden, P. Tapponnier and V. Courtillot for fruitful discussion. P. Tapponnier provided the photographs in Fig. 1.

This is Institute de Physique du Globe contribution no. 578.

Received 6 January; accepted 11 March 1982.

1. Wilson, K. & Kogert, J. *Phys. Rep.* **12C**, 77 (1974).
2. Toulouse, G. & Pfeuty, P. *Introduction au Groupe de Renormalisation et à ses Applications* (Presse Université de Grenoble, 1975).
3. De Gennes, P. G. *Scaling Concepts in Polymer Physics* (Cornell University Press, 1979).
4. Madden, T. R. *Geophysics* **41**, 1104 (1976).
5. Mattauer, M. *Les Déformations de l'Écorce Terrestre*, (Hermann, Paris, 1976).
6. Aki, K. *J. geophys. Res.* **84**, 6140 (1979).
7. Madariaga, R. in *Identification of Seismic Sources*, 71 (Reidel, Dordrecht, 1981).
8. Brace, W. F., Paulding, B. & Scholz, C. H. *J. geophys. Res.* **71**, 3939 (1966).
9. Waversik, W. R. & Brace, W. F. *Rock Mech.* **3**, 61 (1971).
10. Brace, W. F. & Bombolakis, E. G. *J. geophys. Res.* **68**, 3709 (1963).
11. Tapponnier, P. & Brace, W. F. *Int. J. Rock. Mech. Min. Sci. Geomech. Abstr.* **13**, 103 (1976).
12. Scholz, C. H., Sykes, L. R. & Aggarwal, Y. P. *Science* **181**, 803 (1973).
13. Anderson, E. M. *The Dynamics of Faulting and Dyke Formation with Applications to Britain* (Oliver & Boyd, Edinburgh, 2nd ed., 1951).
14. Scholz, C. H. & Kranz, R. *J. geophys. Res.* **79**, 2132 (1974).
15. Hadkey, B. *Proc. Conf. on Tectonic Problems at Low Confining Pressures* (ed. Nur, A.) 427 (CRL, Kovach, 1980).
16. Abe, K. *J. geophys. Res.* **79**, 4393 (1974).
17. Aki, K. *J. geophys. Res.* **73**, 5359 (1968).
18. Mogi, K. *Bull. Earth Res. Inst. Tokyo* **40**, 125 (1962).
19. Kranz, R. L. *J. Rock Mech. Min. Sci. Geomech. Abstr.* **16**, 23 (1979).

Cobalt in north-east Pacific waters

G. A. Knauer, J. H. Martin & R. M. Gordon

Moss Landing Marine Laboratories, Moss Landing, California 95039, USA

Significant understanding has been gained recently about the biogeochemical cycling of trace metals in the ocean. This knowledge has mostly resulted from the accurate measurement of dissolved species in oceanic water columns. We report here that cobalt's vertical distribution is similar to that exhibited[1–3] by Mn; that is, its surface enrichment/deep depletion (Fig. 1). However, amounts of Co (1–7 ng l^{-1}) are ~10–20 times less than those for Mn (Table 1), as might be expected from crustal abundance estimates[4] for these elements (Mn=950; Co = 25 μg per g). The similarity between Mn and Co profiles implies the same biogeochemical pathways. The Co excess in nearshore surface waters probably results from continental weathering input processes, as suggested by the remarkable Co–salinity mirror-image relationship shown in Fig. 1, and the Co–salinity scatter diagram in Fig. 2*a*. The steady decrease in Co concentrations also indicates that Co is usually scavenged rather than regenerated at depth, as is the case with Mn (Fig. 1; Table 1).

Table 1 Dissolved Co levels observed off central California (36° 30′ N 123°05′ W) in June 1981

Depth (m)	OE Co 1 (ng l^{-1})	OE Co 2 (ng l^{-1})	$\bar{X}$ OE Co (ng l^{-1})	CX Co* (ng l^{-1})	CX Mn† (ng l^{-1})
0.2 raft	7.0	6.9	7.0	—	—
10	6.6	6.2	6.4	1.8	130
30	6.8	7.3	7.0	—	160
50	6.6	6.0	6.3	5.1	—
100	5.8	5.9	5.8	4.7	62
200	4.5	3.6	4.0	4.0	40
300	3.9	3.9	3.9	3.3	42
400	3.6	3.4	3.5	2.2	49
520	25.0‡	2.9	2.9	1.1	58
700	2.2	1.8	2.0	2.0	50
980	2.1	1.6	1.8	2.0	45
1,160	1.5	2.1	1.8	2.0	43
1,900	1.8	1.5	1.6	1.9	42
2,350	1.4	1.4	1.4	1.6	31

Cobalt was preconcentrated in two replicates (OE Co 1, OE Co 2) using organic extraction. Replicate values were averaged ($\bar{X}$ OE Co) and are compared with Co amounts obtained using chelex preconcentration (CX Co). Manganese data (CX Mn) for these samples are also shown. Water was filtered using 0.45 μm pore size nucleopores.
* Chelex at pH 6.
† Chelex at pH 8.
‡ Contaminated.

In comparison with those in oceanic waters, Co values can be very high in estuaries. We measured dissolved Co amounts varying between 50 and 1,200 ng l^{-1} in north San Francisco Bay on a Sacramento River/Golden Gate transect during a period of high freshwater flow in March of 1980. The highest concentrations were observed at three stations with salinities between 6.2 and 12.4‰; that is 1,200, 550 and 530 ng l^{-1} (Fig. 2*c*). Cobalt levels immediately outside the above salinity range were markedly lower: 120 and 140 ng l^{-1}, at salinities of 2.2 and 13.8‰, respectively. Amounts of Co more or less steadily decreased in the remainder of the salinity range sampled; that

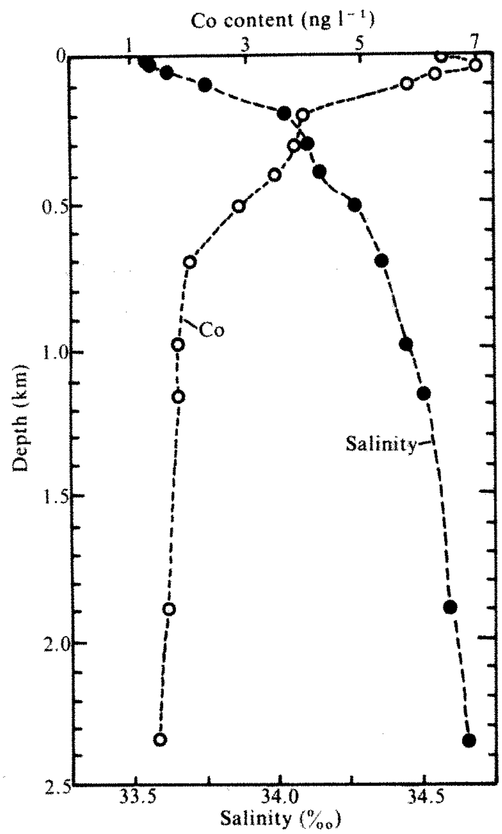


Fig. 1 Salinity (●) and mean organic extraction Co values (○) versus depth (see also Fig. 2*a*).

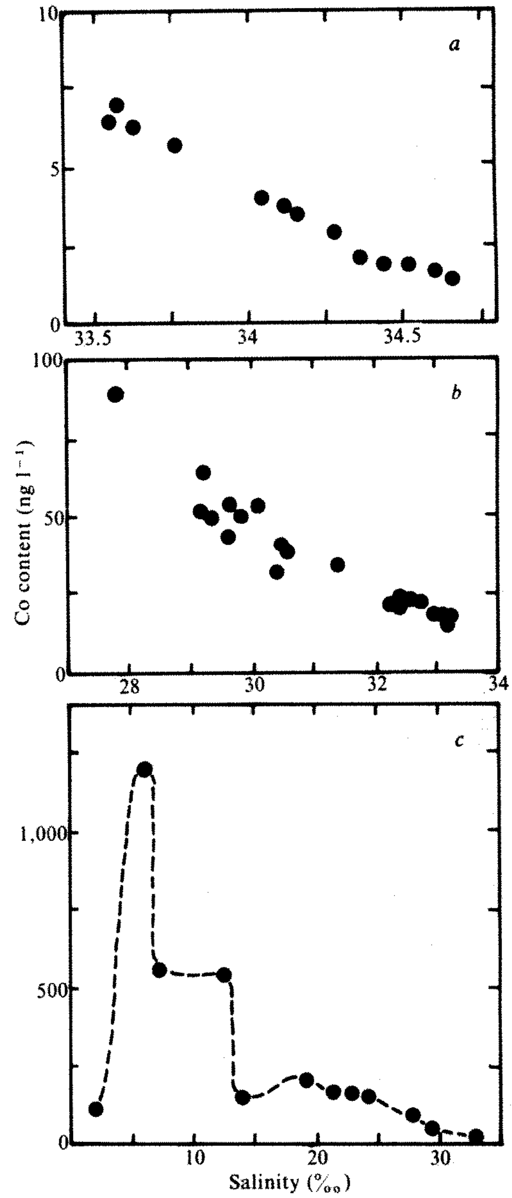


Fig. 2 Cobalt–salinity relationships observed *a*, off central California at depths of 0.2–2,350 m; *b*, in near-surface waters immediately outside San Francisco Bay; *c*, in San Francisco Bay. Correlation coefficients for data in *a* and *b* are −0.93 and −0.99, respectively.

is, from 200 ng l^{-1} at 19.3‰ to 40 ng l^{-1} at 33‰ (Fig. 2*c*). A similar Co–salinity relationship was observed for near-surface waters collected outside San Francisco Bay (Fig. 2*b*). The slight curvature in the plotted points suggests the possibility of Co removal processes occurring within this salinity range, as well as conservative mixing.

The reasons for the very high dissolved Co levels in the San Francisco Bay intense mixing zone (6–12‰ salinity range) waters are not understood. Although the data suggest Co adsorption–desorption reactions occurring with suspended particulates, the exact nature of these reactions is unknown. River–estuarine chemistry is very complicated, involving many competing processes, and is beyond the scope of this paper. Nevertheless, scientists dealing with this complex subject also report similarities between Co/Mn geochemistry[5].

We compared our Co distribution findings with those reported previously[6–13], and in general, values for estuaries and rivers are similar to those we report for near-shore and San Francisco Bay waters, while quantities for various oceanic waters are usually considerably higher. Because most of these data were reported well before the adoption of clean techniques, it is logical to suspect contamination problems for many of the

previously reported oceanic samples, especially those collected at depth.

We avoided contamination by using the same collection, filtration and analytical procedures that we had previously used (with Bruland) for other elements[14]. The organic extraction (APDC, DDDC/chloroform) preconcentration method had yields of 97–98% when checked with carrier-free ^{57}Co, and the precision for two replicates was generally excellent (Table 1). The use of another preconcentration technique (chelex ion exchange at *p*H 6) resulted in similar values (within 1.3 ng) with the exception of the 10 and 520 m samples (Table 1). Analyses of eight individual blanks for the organic extraction procedure resulted in a mean and standard deviation of 0.3 ± 0.04 total ng Co in comparison to 0.6 total ng found in the lowest sample. Cobalt was not detected in the chelex blanks.

In conclusion, the oceanic vertical distribution of dissolved Co appears to be very similar to that of Mn, at least as we now know it. Great variability in near-shore surface waters can be expected and, as is the case with Mn, oceanic removal processes are probably rapid, with resulting residence times in tens of years. This can be illustrated by substituting a world ocean Co concentration of 1 ng l^{-1} in place of the 23 ng l^{-1} used by Bewers and Yeats[15] in their estimates of Co residence times. The results are 52 yr using stream input data and 34 yr using sedimentation rate data.

If the Mn–Co similarity is also true for open-ocean waters, much lower Co levels (~1.0 ng l^{-1}) can also be expected in oligotrophic surface waters, judging by Mn data (~35 ng l^{-1}) reported by Landing and Bruland[2]. Because Co is an important element in many essential biological compounds[16], notably the central atom of vitamin B12, the low amounts of this element in the environment suggest the possibility of its being a limiting nutrient in some conditions, or at least that the massive amounts (5 μg l^{-1}) routinely added to phytoplankton culture media are unnecessary.

We thank S. Fitzwater, S. Tanner and C. Hunter and the captain and crew of R/V *Cayuse* for their help during our Cobalt I cruise. This research was supported by grants from the NSF Biological Oceanography (OCE 80-03200) and Marine Chemistry (OCE 79-09431) programmes, the Ocean Minerals Corporation, and the US Environmental Protection Agency (CR 8077110020).

Received 22 December 1981; accepted 15 March 1982.

1. Martin, J. H. & Knauer, G. A. *Earth planet. Sci. Lett.* **51,** 266–274 (1980).
2. Landing, W. M. & Bruland, K. W. *Earth planet. Sci. Lett.* **49,** 45–56 (1980).
3. Klinkhammer, G. P. & Bender, M. L. *Earth planet. Sci. Lett.* **46,** 361–384 (1980).
4. Taylor, S. R. *Geochim. cosmochim. Acta* **28,** 1273–1285 (1964).
5. Sholkovitz, E. R. & Copland, D. *Geochim. cosmochim. Acta* **45,** 181–189 (1981).
6. Kharkar, D. P., Turekian, K. K. & Bertine, K. K. *Geochim. cosmochim. Acta* **32,** 285–298 (1968).
7. Weiss, H. V. & Reed, J. A. *J. mar. Res.* **18,** 185–188 (1960).
8. Thompson, T. G. & Laevastu, T. *J. mar. Res.* **18,** 189–193 (1960).
9. Schutz, D. F. & Turekian, K. K. *J. geophys. Res.* **70,** 5519–5528 (1965).
10. Robertson, D. E. *Geochim. cosmochim. Acta* **34,** 553–567 (1970).
11. Spencer, D. W., Robertson, D. E., Turekian, K. K. & Folsom, T. R. *J. geophys. Res.* **75,** 7688–7696 (1970).
12. Windom, H. L. & Smith, R. G. *Deep-Sea Res.* **19,** 727–730 (1972).
13. Bewers, J. M., Sundby, B. & Yeats, P. A. *Geochim. cosmochim. Acta* **40,** 687–696 (1976).
14. Bruland, K. W., Franks, R. P., Knauer, G. A. & Martin, J. H. *Analyt. chim. Acta* **105,** 233–245 (1979).
15. Bewers, J. M. & Yeats, P. A. *Nature* **268,** 595–598 (1977).
16. Bowen, H. J. M. *Trace Elements in Biochemistry* (Academic, London, 1966).

Beryllium in the water column of the central North Pacific

C.I. Measures & J. M. Edmond

Department of Earth and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

The existence of the cosmogenic radioisotope of beryllium, ^{10}Be($t_{1/2}$ = 1.5 Myr), was first confirmed[1,2] almost 25 yr ago. Subsequently, Merrill *et al.*[3] worked out the sedimentary geochemistry of the stable isotope of the element, ^{9}Be. The geochronological potential of ^{10}Be could not be exploited, however, because of its extremely low abundance (global average production rate 1.08 atoms cm^{-2} min^{-1})[4] and the difficulties associated with the analysis of the stable species ^{9}Be, particularly in natural waters. This situation is changing dramatically with the application of accelerator techniques to the detection of the radioisotope[5]. It thus becomes important to obtain a detailed understanding of the geochemistry of ^{9}Be since the potential use of the stable isotope for normalization purposes would significantly enhance the usefulness of the radioactive species in dating studies. We report here the first detailed profile for ^{9}Be in the oceanic water column.

Samples were obtained at 30 °N, 158 °W over MANOP site R (Manganese Nodule Program Red Clay Study Area) north of Hawaii using clean oceanographic procedures. One litre samples were collected in acid-cleaned linear polyethylene containers and immediately spiked with 4 ml of doubly distilled 6M HCl to bring the *p*H to ~1.8. Laboratory experiments with standards indicate that there is no loss of beryllium in these storage conditions. The analyses were made by gas chromatography with electron capture detection of the volatile trifluoroacetylacetonate complex of the element. This was formed in the seawater solution at *p*H 5.4 and extracted into high-purity benzene (details of the method will be reported elsewhere). The detection limit using a 200 ml sample aliquot is about 2 pmol kg^{-1} (2×10^{-12} M); the precision (1σ) is 5.5% at 20 pmol kg^{-1}.

The profile is presented in Fig. 1 together with the associated hydrographic measurements. The numerical data are listed in Table 1. The distribution of beryllium in the water column displays the nutrient-like surface depletion and deep enrichment observed for many of the other accurately determined

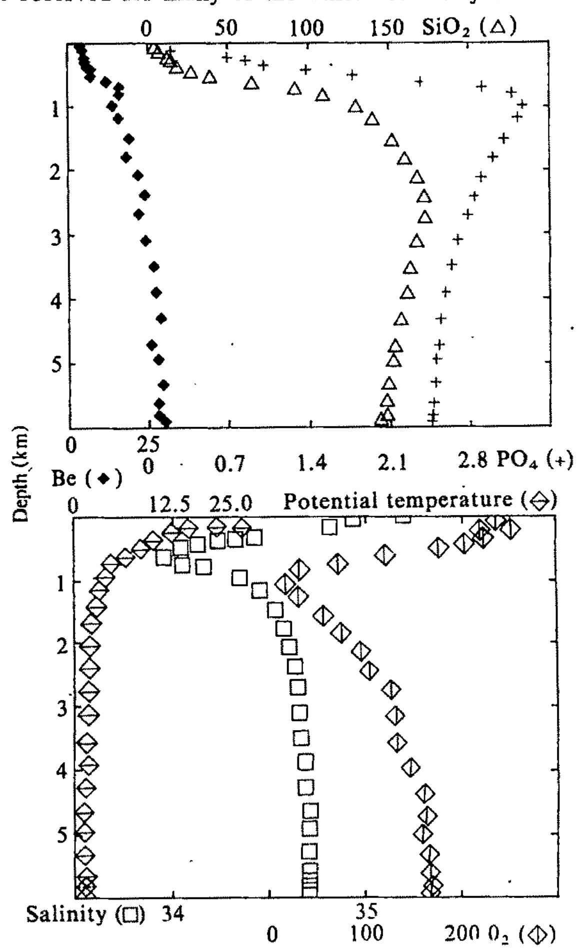


Fig. 1 The beryllium profile from MANOP R in the Central North Pacific (30° N, 158° W). Also shown are the distributions of silicate, phosphate, potential temperature, salinity and oxygen.

0028-0836/82/180051—03$01.00

Table 1 Beryllium and ancillary data from MANOP R (30° N, 156° W)

Depth (m)	Potential temperature (°C)	%S	O_2 (μmol kg^{-1})	PO_4 (μmol kg^{-1})	Si (μmol kg^{-1})	Be (pmol kg^{-1})
40	24.64	35.209	244	0.06	2.6	4.0
70	21.60	34.939	253	0.08	3.8	5.0
110	18.72	34.826	241	0.22	5.0	4.0
206	14.67	34.407	218	0.71	10.8	5.0
266	12.68	34.330	219	0.87	13.6	5.4
336	11.30	34.255	221	1.02	17.5	6.0
417	9.96	34.126	206	1.37	26.1	7.0
497	8.04	34.050	178	1.78	38.4	7.0
593	6.15	33.985	122	2.37	64.0	12.0
691	4.75	34.058	74	2.91	90.9	16.0
789	4.30	34.179	32	3.17	108.9	16.0
986	3.67	34.364	20	3.26	129.2	14.0
1,182	3.15	34.462	30	3.22	139.5	16.0
1,505	2.53	34.542	57	3.10	151.3	19.0
1,804	2.17	34.588	76	3.00	159.2	18.0
2,102	1.92	34.618	91	2.90	166.7	22.0
2,401	1.72	34.640	103	2.84	170.5	24.0
2,699	1.61	34.651	126	2.78	171.0	22.0
3,101	1.32	34.665	130	2.69	165.3	24.0
3,508	1.22	34.671	132	2.63	161.7	26.0
3,915	1.16	34.679	147	2.58	159.7	27.0
4,322	1.12	34.684	162	2.55	155.8	28.0
4,730	1.09	34.687	166	2.52	151.7	25.0
4,921	1.08	34.686	162	2.50	150.4	28.0
5,314	1.07	34.688	167	2.49	147.7	29.0
5,608	1.05	34.689	166	2.48	146.7	28.0
5,805	1.05	34.690	168	2.47	146.1	28.0
5,903	1.05	34.689	167	2.46	145.1	30.0

trace elements. The mixed layer value is 4 pmol kg^{-1}; the concentrations rise through the main thermocline to values of 25–30 pmol kg^{-1} in the deep and bottom waters. Although there may be a secondary maximum associated with the oxygen minimum at 900 m (Fig. 1), the sampling density is inadequate to resolve it fully.

MANOP R is located beneath the tongue of Antarctic Bottom Water that penetrates north around Hawaii[6]. It is the hydrographic feature responsible for the induced maximum in the silica profile that characterizes the North Pacific Deep Water (Fig. 1). The beryllium profile, in contrast to those of the nutrients, increases with depth. The potential temperature–

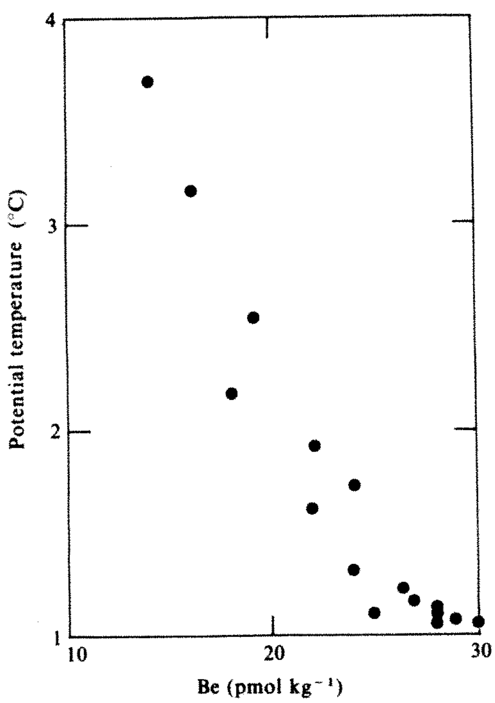


Fig. 2 The beryllium–potential temperature relationship for the deep and bottom waters.

concentration relationship below 1,000 m (Fig. 2) is concave-up indicating mid-water scavenging[7] as has been observed for copper[8]. The fit to a simple one-dimensional diffusion advection model[7] gives a value for the ratio of the scavenging rate, J, to the vertical advection velocity, w, of $J/w = 2.49 \times 10^{-3}$ pmol kg^{-1}yr^{-1}m^{-1} with a standard error of 1.90. Assuming that the upwelling velocity is 3 m yr^{-1} (ref. 7) then the scavenging rate, J, is 7.5×10^{-3} pmol kg^{-1}yr^{-1}. The half life with respect to scavenging is thus estimated to be 1,850 yr. As the concentration of beryllium increases with depth there must be a source of input from the bottom to maintain the profile against removal by scavenging. This can be crudely estimated[9] at 1.5 pmol cm^{-2}yr^{-1}. Assuming an effective pore water diffusion coefficient of 10^{-6} cm^{-2} s^{-1} then this flux requires a gradient in the sediments of ~60 pmol kg^{-1} cm^{-1}. As in the case of copper, such high gradients cannot extend very deep into the sediments because the solubility of any existing beryllium compounds will be exceeded[10]. Thus a strong source is required near the surface.

Beryllium exists in the oceans at concentrations as low as those reported for any other trace element (see Fig. 3). The levels in rivers, however, are substantially higher. Merrill *et al.*[3] reported values of a few nmol kg^{-1}(10^{-9}M), which we have confirmed. Three samples from the main channel of the lower Amazon give values of 1.6, 1.7 and 1.8 nmol kg^{-1}, while an estimate from the mouth of the Orinoco is 610 pmol kg^{-1}. The Yangtze gives lower values: 140 pmol kg^{-1} at the mouth and 150 pmol kg^{-1} at Wuhan in the central part of the flood plain. If the global average value is taken as 1 nmol kg^{-1} then the fluvial flux is 33×10^6 mol yr^{-1}. Assuming that the element behaves conservatively during estuarine mixing and that the average oceanic concentration is 20 pmol kg^{-1}, then the residence time for the element is about 800 yr. As this is shorter than the oceanic turnover time it is important that the value be confirmed in future work.

Crucial to the application of ^{10}Be in marine geochronology is the degree of uniformity of the ^{10}Be/^{9}Be ratio in the oceanic water column. If tracer equilibrium is achieved during oceanic mixing then changes in the initial uptake of the radioisotope in the sedimentary substrate caused by changes in, for example, composition and accumulation rate can be corrected for directly. The concentration of ^{10}Be in seawater has been

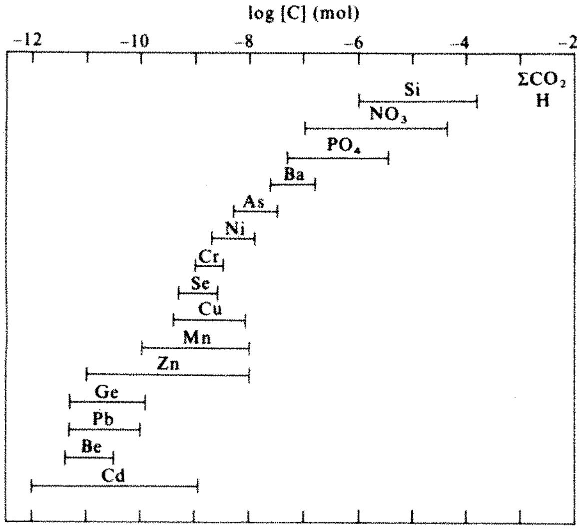


Fig. 3 Published concentration [C] ranges for oceanic trace metals. The nutrients and CO_2 are shown for comparison. Data sources: Ba, ref. 15; As, ref. 16; Cu, Ni, Cd and Zn, ref. 17; Cr, ref. 18; Se, ref. 19; Mn, ref. 20; Ge, ref. 21; Pb, ref. 22.

measured using accelerator techniques[11,12]. An average value for nutrient-depleted Pacific surface waters (two samples) is $1.2 \pm 0.3 \times 10^{-18}$ mol kg^{-1} (ref. 11). On the assumption of a ^{9}Be concentration of 4 pmol kg^{-1}, the ^{10}Be/^{9}Be ratio becomes $300 \pm 75 \times 10^{-9}$. A ^{10}Be concentration of $10.2 \pm 2 \times 10^{-18}$ mol kg^{-1} reported for a single sample from a depth of 4,100 m south-east of San Diego[12], coupled with the deep-water ^{9}Be value of 27 pmol kg^{-1} at MANOP R, give an isotope ratio of $377 \pm 75 \times 10^{-9}$, which is consistent with the surface value.

Yokoyama *et al.*[13] and Krishnaswami *et al.*[12] have reported isotope ratios in manganese nodules (extrapolated to zero age) of 85×10^{-9} and 114–$1{,}215 \times 10^{-9}$ with a mean of 255×10^{-9} respectively. Given the present accuracy of the ^{10}Be measurements ($\pm 20\%$)[12,14] the overall similarity of the water column and nodule data suggests that the two isotopes are indeed equilibrated in the oceans. Much more work is needed, however, before this important conclusion can be taken as firmly established. In particular, the two isotopes need to be measured on the same water samples. In addition, leaching experiments must be carried out on sediments and nodules to establish that only authigenic beryllium is being analysed rather than an additional component of 'dead' detrital beryllium.

We acknowledge the support and encouragement of our colleagues in the MANOP Program of the NSF/IDOE. We also thank the Office of Naval Research for general support of our analytical program, and Professor K. K. Turekian for helpful discussions.

Received 29 December 1981; accepted 18 February 1982.

1. Arnold, J. R. *Science* **124,** 584 (1956).
2. Goel, P. S. *et al. Deep-Sea Res.* **4,** 202 (1957).
3. Merrill, J. R., Lyden, E. F. X., Honda, M. & Arnold, J. R. *Geochim. cosmochim. Acta* **18,** 108 (1960).
4. Reyss, J. I., Yokoyama, U. & Guichard, F. *Earth Planet. Sci. Lett.* **53,** 203 (1981).
5. Raisbeck, G. M., Yiou, F., Fruneau, M. & Loiseaux, J. M. *Science* **202,** 215 (1978).
6. Edmond, J. M., Chung, Y. & Sclater, J. G. *J. geophys. Res.* **76,** 8089 (1971).
7. Craig, H. *Earth planet. Sci. Lett.* **23,** 149 (1974).
8. Boyle, E. A., Sclater, F. R. & Edmond, J. M. *Earth planet. Sci. Lett.* **37,** 38 (1977).
9. Munk, W. *Deep-Sea Res.* **13,** 707 (1966).
10. Everest, A. D. *Comprehensive Inorganic Chemistry* Vol. 1, Ch. 9 (Pergamon, Oxford, 1973).
11. Raisbeck, G. M., Yiou, F., Fruneau, M. & Loiseaux, J. M. *Earth planet. Sci. Lett.* **43,** 237 (1979).
12. Krishnaswami, S. *et al. Earth planet. Sci. Lett.* (in the press).
13. Yokoyama, Y., Guichard, F., Reyss, J. L. & Nguyen, H. V. *Science* **201,** 1016 (1978).
14. Guichard, F., Reyss, J. L. & Yokoyama, Y. *Nature* **272,** 155 (1978).
15. Bacon, M. P. & Edmond, J. M. *Earth planet. Sci. Lett.* **16,** 66 (1972).
16. Andreae, M. O. *Limnol. Oceanogr.* **24,** 440 (1979).
17. Bruland, K. W. *Earth planet. Sci. Lett.* **47,** 176 (1980).
18. Cranston, R. E. & Murray, J. M. *Analyt. Chim. Acta* **99,** 275 (1978).
19. Measures, C. I., McDuff, R. E. & Edmond, J. M. *Earth planet. Sci. Lett.* **49,** 102 (1980).
20. Landing, W. M. & Bruland, K. W. *Earth planet. Sci. Lett.* **49,** 45 (1980).
21. Froelich, P. N. & Andreae, M. O. *Science* **213,** 205 (1981).
22. Schaule, B. K. & Patterson, C. C. *Earth planet. Sci. Lett.* **54,** 97 (1981).

Hindered bedload settling as a model of sand bed planation by water waves

Robert J. Hallermeier

Coastal Engineering Research Center, Fort Belvoir, Virginia 22060, USA

The interrelationships between fluid flows and the surface forms of underlying movable beds are crucial in interpreting sedimentary structures and in predicting hydraulic drag. Energetic flows can erode all features from a sediment bed, and transition to such a planar bed is important to the processes of sediment transport by waves[1]. The following analysis connects bed planation to a threshold effect in hindered settling with increasing concentration of noncohesive sediment moving near the bed. Calculated fluid velocity from the resulting quantitative criterion for this bed transition agrees with extensive laboratory data[2,3] in oscillatory flows.

The present model of the planation process utilizes simple empirical forms of several of the relationships that are needed. As Reynolds numbers for bed planation are usually large enough for the near-bed flow to be intensely mixed rather than laminar[4], inertial (non-viscous) asymptotes are suitable for most elements of the model. However, fully-developed turbulent flow is absent in the laboratory situations and Jonsson[5] has pointed out the persistent empirical accuracy through this transitional range of one result [equation (4)] from laminar flow.

The primary element in the model is an expression[6] for transport rate of sediment moving near the bed in oscillatory flow:

$$Q/\omega D^2 = [\Phi/10]^{1.5} \tag{1}$$

where Q is average bedload sediment volume per unit time and bed width (moving to and fro colinear with the flow), and $\Phi = [(a\omega)^2/\gamma' gD]$ measures sediment agitation by flow energy. Here ω is flow (wave) frequency, D is median sediment diameter, $a\omega$ is peak near-bed horizontal velocity, g is gravitational acceleration and $\gamma' = (\rho_s/\rho - 1)$, where ρ_s is sediment density and ρ is fluid density. Note that bedload rates reported by Manohar[2] agree with equation (1) in a range of conditions through the transition from rippled to planar bed[6].

Conversion of the bedload rate Q into volumetric transport concentration, the ratio of sediment to fluid discharge, requires

Table 1 Test sediments represented in Fig. 1

Label	Sediment description γ'	D(mm)	A	n	Maximum $(a\omega^2/g)$ included
Manohar[2] spherical grains					
a	1.49	0.235	234	4.2	0.18
b	1.54	0.61	4,590	3.1	0.25
Manohar[2] other grains					
c	1.65	0.28	420	3.65	0.22
d	1.63	0.786	9,660	2.3	0.32
e	1.60	1.006	18,900	2.3	0.32
f	0.28	3.17	101,000	2.3	0.09
g	1.60	1.829	108,000	2.3	0.31
h	1.63	1.983	137,000	2.3	0.32
Chan *et al.*[3] spherical grains					
j	1.65	0.097	18.7	4.6	
k	0.32	0.359	74.4	4.5	
l	1.65	0.254	335	4.05	
Chan *et al.*[3] other grains					
m	0.97	0.254	79.5	4.4	
n	1.55	0.254	315	3.8	
p	0.97	0.505	625	3.4	
Kennedy and Falcon[18] spherical grains					
q	0.035	1.0	276	4.1	

0028-0836/82/180053—03$01.00

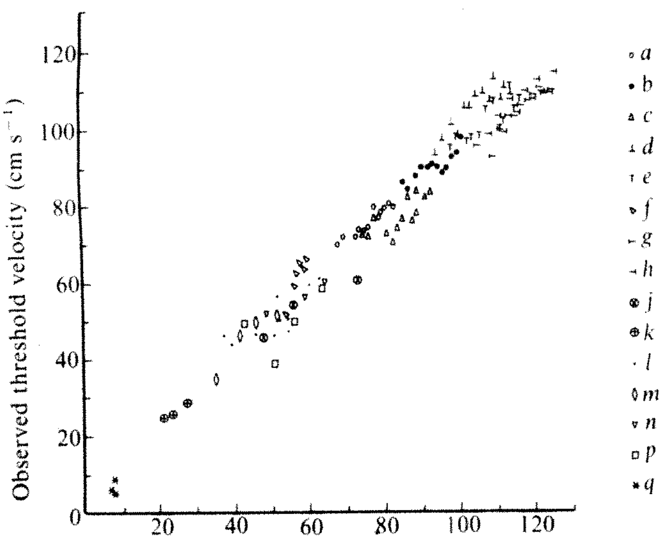


Fig. 1 Threshold velocities for bed planation observed in 120 laboratory tests[2,3,18] compared with calculations using equation (11). See Table 1 for description of the 15 sediments represented.

division by Λ, a characteristic thickness of the bedload layer, and by Ψ, a characteristic fluid velocity within the layer. Bagnold[7] reported that the elevation above the bed of this layer's centre (in unidirectional flow) is mD, where

$$m = 1.4(u_*/u_{*0})^{0.6} \quad (2)$$

and the estimate adopted here is $\Lambda = 2mD$. Here u_* is the friction velocity and u_{*0} is critical friction velocity for sediment motion initiation (zero transport). The peak friction velocity is related to peak velocity by[5]

$$u_* = a\omega(f_w/2)^{0.5} \quad (3)$$

where f_w is friction factor, depending on boundary roughness and on the Reynolds number $(a^2\omega/\nu)$, ν being fluid viscosity. For the situations of interest with a nearly plane bed, boundary-layer flow is transitional between laminar and turbulent[8], but the analytical form for a smooth boundary

$$f_w = 2/(a^2\omega/\nu)^{0.5} \quad (4)$$

remains empirically valid although derived for laminar conditions[5]. The proper inertial relation for sand motion initiation in oscillatory flow is[9]

$$(a\omega)_0 = (8\gamma' gD)^{0.5} \quad (5)$$

so that

$$u_{*0} = (a\omega)_0^{0.5}(\nu\omega)^{0.25} = (8\gamma' gD\nu\omega)^{0.25} \quad (6)$$

A suitable velocity profile for estimating Ψ is problematical even in steady transitional (intermittently turbulent) flow[10]. The fairly thick bedload layer would appear to be dominated by peak-flow effects here, so the appropriate approximation seems to be an inertial profile with logarithmic variation of velocity, as occurs beyond the viscous sublayer for fluid flow above moderately large Reynolds number [10]. The mean horizontal fluid velocity is taken as[7]

$$u = 5.75\, u_* \log(30z/k) \quad (7)$$

where z is distance from the boundary and k measures boundary roughness. For a nearly plane bed $k = 2.5\,D$ is appropriate[8,11], and the characteristic velocity is[7] that at elevation $z = 0.37\,\Lambda$. With these relationships

$$\Psi = 5.75\,(a\omega)^{0.5}(\omega\nu)^{0.25}\log[12.4(u_*/u_{*0})^{0.6}] \quad (8)$$

The postulated link between bed planation and hindered sediment settling is pursued by regarding the right side of equation (1). The agitation index Φ has[12] an inverse dependence on the square of inertial sediment settling velocity, so that equation (1) indicates an inverse dependence between bedload rate and the cube of settling velocity[6]. The terminal settling velocity, w, of an isolated sphere in still fluid is described[13] as a variation of (wD/ν) with the buoyancy index $A = (\gamma' gD^3/\nu^2)$, having asymptotic behaviours of viscous settling for $A < 20$ and inertial settling for $A > 10^5$. A homogeneous dispersion of cohesionless spheres settles more slowly than an isolated sphere, with velocity given as[14]

$$w_C = w(1-C)^n \quad (9)$$

where C is volumetric particle concentration, and the exponent n has a smooth variation between $n = 4.6$ for viscous settling and $n = 2.3$ for inertial settling. An isolated natural sediment grain usually settles[15] at slightly smaller velocity than a sphere of the same mass due to increased drag, and inertial settling arises for $A > 10^4$. Effects with natural grain dispersions presumably follow equation (9) if adjustment is made for n between more closely spaced asymptotes.

No data are available for entrained sediment concentrations near a planar bed, but a critical concentration for appreciably hindered settling apparently should be related to grain interactions and the threshold for significant grain–grain collisions has been reported[16,17] to be C near 0.09 or 0.02. Here $C_{cr} = 0.075$ is adopted; this value is one-eighth of 0.6, the usual maximum possible sediment concentration. Thus the postulated threshold for bed planation becomes

$$(Q/\Lambda\Psi)_{cr} = C_{cr} = \omega D^2(\Phi/10)^{1.5}(1-C_{cr})^{-3n}(\Lambda\Psi)^{-1} \quad (10)$$

or

$$(a\omega)^{2.2} = 28\,\frac{(0.925)^{3n}\nu^{0.25}(\gamma' g)^{1.35}D^{0.35}}{\omega^{0.75}}\log\left[\frac{12.4(a\omega)^{0.3}}{(8\gamma' gD)^{0.15}}\right] \quad (11)$$

With given material characteristics, equation (11) is an implicit relationship for flow conditions at the postulated threshold of bed planation: specifying either flow frequency (ω) or near-bed horizontal flow amplitude (a), the critical value of the other variable is determinable by root-finding. To clarify the dependences of the above threshold, an approximate explicit form is

$$a\omega = 2.8(\gamma' g)^{0.64}D^{0.22}\nu^{0.07}\omega^{-0.35} \quad (12)$$

For the test conditions considered here, equation (12) regularly gives a threshold velocity within a few per cent of that determined by the root of equation (11).

Calculated threshold velocities using equation (11) are compared in Fig. 1 with results from 120 laboratory tests of bed planation with the sediments described in Table 1. Calculations are close to the (very low) near-bed threshold velocities reported for three wave-tank tests conducted by Kennedy and Falcon[18] at transition from rippled to planar bed with lightweight spheres. However, only part of the experimental results from the two major oscillatory-flow studies[2,3] are presented in Fig. 1 for the following reasons.

Manohar[2] performed tests by oscillating at fixed amplitudes a sediment tray in still water, and observing the critical frequency for bed planation; analogously, equation (11) has been solved with the specified a to find the threshold ω. Ignoring the independent ω effects, Manohar provided this correlation of results:

$$a\omega = 20.9(\gamma' g)^{0.4}\nu^{0.2}D^{0.2} \quad (13)$$

For accelerated-bed tests, Bagnold[19] reported that general sliding of sediment can occur at large accelerations, depending on ρ_s/ρ and the angle of sediment repose. Consequently a minority of these tests with large $a\omega^2$ has been disregarded, with the cutoff values of $a\omega^2/g$ listed in Table 1; Manohar's tabulated results at small amplitude increments exhibit in each test series a sizeable gap in $a\omega^2$ and $a\omega$ around the stated $a\omega^2/g$. For the remaining 90 tests, with appropriate values of n (Table 1),

agreement is evident between experimental and calculated results, although there is a tendency with some sediments for observed threshold velocity to be ~10% less than that calculated.

Calculations show agreement only with a quite limited segment of results from the other major study. Chan *et al.*[3] executed tests by oscillating fluid at fixed frequencies above stationary sediment in a 5-cm diameter pipe, observing critical fluid amplitude for bed planation; equation (11) has been solved by a corresponding procedure. About one-third of the material combinations tested are represented in Fig. 1; these correspond to A values for subaqueous fine and medium quartz sands (although some results for both water and ethanol are shown). More scatter is to be expected here because results have been scaled off a log-log plot for the present use. The agreement of these 27 measurements with calculations by equation (11) is about as ideal as with transition calculations by the following expression, provided as an empirical correlation of all results by Chan *et al.*[3]:

$$a\omega = 6.6(\gamma' g)^{0.5}\nu^{0.2}D^{0.1}\omega^{-0.2} \tag{14}$$

On the other hand, equation (14) is clearly superior to equation (11) in agreement with the data for other sets of tests with: large D, large ν (glucose) or iron ore. The causes for the present model's failure are not certain, but these cases might be explained by the facts that: equation (4) is inaccurate for relatively large D/a (ref. 8); the small test facility gives[20] laminar flow of glucose, making equation (7) inapplicable; and iron ore exhibits cohesion by its magnetism, making equation (9) incorrect. Also, no heavy minerals such as iron ore are represented in the data base[6] supporting equation (1).

I conclude that the model of bed planation providing equation (11) is consistent with laboratory results for 15 sediments having A values corresponding to quartz sand in water, and segments of the two major data sets[2,3] show some congruence, whereas these results had appeared disjoint[3,21]. Note that this analysis has a markedly different viewpoint from that of Bagnold's criterion for planar bed occurrence[7,22], and that the latter has been supported in data reviews[1,23] for unidirectional and oscillatory flows. The present results are established for laboratory transitions to a planar bed from ripples with wavelength governed by a, rather than from bedforms occurring in flows at lower frequencies[21].

The results reported here, unless otherwise noted, are based on research conducted at the Coastal Engineering Research Center, US Army Corps of Engineers. The findings are not to be construed as an official US Department of the Army position unless so designated by other authorized documents. Permission to publish the present results is appreciated.

Received 7 December 1981; accepted 19 March 1982.

1. Komar, P. D. & Miller, M. C. *J. sedim. Petrol.* **45**, 697–703 (1975).
2. Manohar, M. *Beach Eros. Bd Tech. Mem.* No. 75 (1955).
3. Chan, K. W. *et al. Proc. R. Soc.* **A330**, 537–559 (1972).
4. Sleath, J. F. A. *J. WatWays Harb. coast. Engng Div. Am. Soc. civ. Engrs* **100** (WW2) 105–122 (1974).
5. Jonsson, I. G. *ISVA Ser. Pap.* **17** (Technical University of Denmark, Lyngby, 1978).
6. Hallermeier, R. J. *Continental Shelf Res.* (submitted).
7. Bagnold, R. A. *Proc. R. Soc.* **A332**, 473–504 (1973).
8. Kamphuis, J. W. *J. WatWays Harb. coast. Engng Div. Am. Soc. civ. Engrs* **101** (WW2) 135–144 (1975).
9. Hallermeier, R. J. *J. WatWay Port coast. Ocean Div. Am. Soc. civ. Engrs* **106** (WW3) 299–318 (1980).
10. Schlichting, H. *Boundary-Layer Theory*, 6th edn (McGraw-Hill, New York, 1968).
11. Engelund, F. & Hansen, E. *A Monograph on Sediment Transport in Alluvial Streams* (Teknisk Forlag, Copenhagen, 1967).
12. Nayak, I. V. *Tech. Rep. HEL 2-25* (University of California, Berkeley, 1970).
13. Yalin, M. S. *Mechanics of Sediment Transport* 2nd edn (Pergamon, Oxford, 1977).
14. Richardson, J. F. & Jeronimo, M. A. daS. *Chem. engng Sci.* **34**, 1419–1422 (1979).
15. Hallermeier, R. J. *Sedimentology* **28**, 859–865 (1981).
16. Bagnold, R. A. *Phil. Trans. R. Soc.* **A249**, 234–297 (1956).
17. Leeder, M. R. *Earth Surf. Processes* **4**, 229–240 (1979).
18. Kennedy, J. F. & Falcon, M. *Hydrodyn. Lab. Rep.* **86** (Massachusetts Institute of Technology, Cambridge, 1965).
19. Bagnold, R. A. *Proc. R. Soc.* **A187**, 1–18 (1946).
20. Hino, M. *et al. J. Fluid Mech.* **75**, 193–207 (1976).
21. Lofquist, K. E. B. *Coast. Engng Res. Cen. Tech. Pap.* **78-5** (Fort Belvoir, Virginia, 1978).
22. Bagnold, R. A. *Geol. Surv. Prof. Pap.* **422-I** (US Printing Office, Washington DC, 1966).
23. Allen, J. R. L. & Leeder, M. R. *Sedimentology* **27**, 209–217 (1980).

Acid rain on Bermuda

Timothy Jickells*, Anthony Knap*, Thomas Church†, James Galloway‡ & John Miller§

* Bermuda Biological Station, Ferry Reach 1-15, Bermuda
† College of Marine Science, University of Delaware, Newark, Delaware 19711, USA
‡ Department of Environmental Science, University of Virginia, Charlottesville, Virginia 22903, USA
§ Air Resources Laboratory, NOAA, Washington DC 20910, USA

Increased acidity of precipitation due to combustion of fossil fuels has been well documented for both the eastern USA[1] and Canada[2]. The SO_2 and NO_x emitted by the burning of coal, natural gas, fuel oil and petrol are oxidized in the atmosphere to sulphuric and nitric acids which subsequently give rise to acid precipitation[1]. However, the SO_2 and NO_x emitted, and their oxidation products, are not all removed by atmospheric deposition over the North American continent; a large fraction is advected east out of North America[3]. In a study between 1 May 1980 and 30 April 1981, we have detected acid precipitation (pH < 5.6) on the island of Bermuda, which is ~1,000 km east of the Atlantic seaboard of the USA. We report here that the acidity of such precipitation is eight times greater on a volume-weighted annual average than rainwater in natural atmospheric equilibrium, and that the acids present are almost wholly sulphuric with a small nitric acid contribution. There is a strong correlation between the presence of these strong acids and the meteorological back trajectory of Bermuda storm systems to the North American continent, which suggests long-range atmospheric transport of acid rain precursors to Bermuda, these having similar anthropogenic origins (that is, from remote fossil fuel combustion) to acid rain precursors on the continent.

Preliminary studies of rainwater in Bermuda from February 1979 to April 1980 suggested the existence of acid rain in Bermuda with a possible seasonal element[4]. To document these observations further, we have sampled all rainfalls in Bermuda from 1 May 1980 at a site at the eastern end of the island, 60 m above sea level, which is thought to be remote from local pollution. We discuss here the data collected up to 30 April 1981.

Samples were collected on an event basis using a wet-only HASL collector[5] for all months except from late August to mid-September 1980 when the collector malfunctioned. During this period we used a bulk collector manned on an event basis. During the year there were 109 rainfalls for which we measured the pH. In addition, 70 of these rainfalls provided sufficient samples to determine concentrations of all the major anions and cations. Total rainfall during the year was 150.22 cm compared with a long-term average of 146.46 cm (ref. 6). Samples were analysed for pH, strong and total acidity[7] and all major anions and cations, using standard colorimetric, atomic absorption and ion chromatographic techniques. In all cases the analyses and pH measurements are accurate to better than 10% and ±0.1 pH units, respectively. The results for all wet-only samples are presented in Table 1.

We have found that precipitation in Bermuda is acidic (Fig. 1): For the entire sampling period, the volume-weighted mean pH of all wet-only samples was 4.74. Gran's titrations[7] were performed on 29 samples having a pH < 4.8. Regression analysis of free acidity, derived from the pH (measured by glass electrode), against strong acidity (determined by Gran's titration) for these 29 samples gave a slope of 0.95 and a regression coefficient, r, of 0.947, indicating that strong acids are responsible for more than 90% of the acidity in Bermuda's rainwater.

The results in Table 1 indicate the presence of excess calcium, potassium, sulphate, nitrate and ammonia over that introduced to the rainwater by sea salts. The calculations of excess are

0028-0836/82/180055—03$01.00

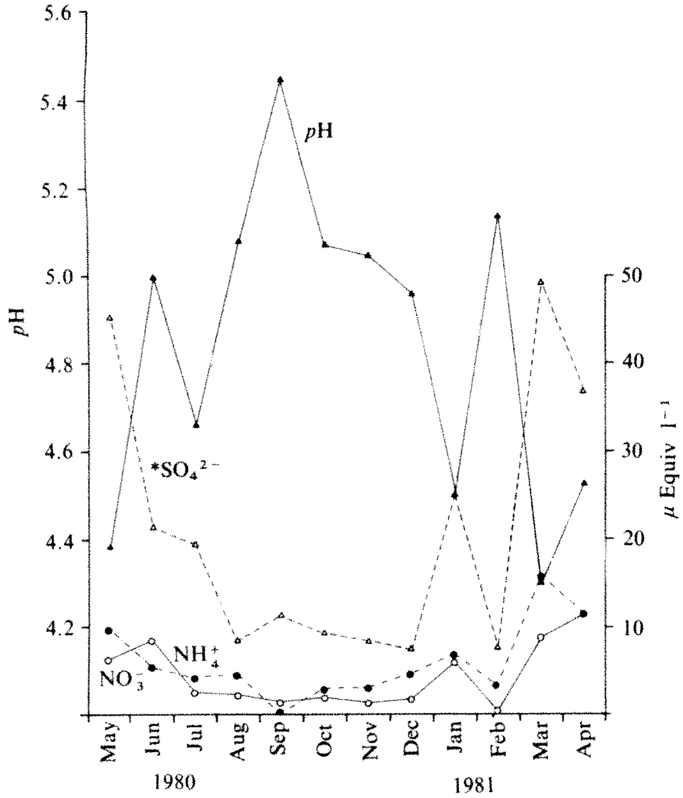


Fig. 1 Volume weighted monthly mean *p*H (▲) and concentration of NO_3^- (●), $^*SO_4^{2-}$ (△) and NH_4^+ (○) of rainfall collected in wet-only samples from Bermuda. * Non-sea salt component. The accuracy of each individual analysis is better than 10% and *p*H measurements are accurate to better than ±0.1 *p*H units.

based on sodium, which is assumed to be derived entirely from unfractionated sea salt. The source of the excess calcium and potassium is believed to be clay and rock particulates of either remote aeolian or local limestone origins. The excess sulphate and nitrate in the rainwater is responsible for the acidity. The amount of basic unfractionated sea salt (bicarbonate and boric acid) is insufficient to cause significant neutralization of the acidity. When protons are correlated with non-sea salt sulphate and nitrate concentrations (12 data points per parameter) the slopes are 0.935 ($r = 0.941$) and 0.257 ($r = 0.909$), respectively. This is interpreted to mean that the upper limits of H_2SO_4 and HNO_3 contributions to free acidity are 94% and 26%, respectively.

There are three possible sources of the sulphuric acid in Bermuda rain: locally derived acids, those produced by natural marine processes, or long-range transport of continentally-derived acids. A locally-derived source of the acids is unlikely to be important because: (1) there is no manufacturing industry in Bermuda; (2) the sampling site was 15 km from the island's only power generating plant and during most storms the site was to windward of the plant. (3) Although the sampling site was near the airport, we do not believe that this affected the precipitation acidity because the acidity is primarily due to H_2SO_4, whereas aircraft exhaust emits mainly NO_x relative to SO_2 and SO_4 (ref. 8). (4) The air mass transit time across the island is very short, giving rapid dispersion and dilution of locally derived pollutants. The meteorology of Bermuda exhibits little or no 'island effect' (ref. 9 and M. J. Nemkosky, personal communication). (5) The observed seasonal pattern described below is incompatible with a local source as power use and air traffic are at a maximum in summer in Bermuda while winds are at a minimum. These factors would produce a summer *p*H minimum if local sources were important, which is the reverse of the pattern observed.

The next possibility is that of natural marine processes giving rise to acidic gases. The emission of various volatile sulphur compounds from seawater has been reported elsewhere[10]. We do not have a direct measure of the significance of these processes, but using the data of Nguyen *et al.*[10], it is possible to make a rough calculation of their significance; the total potential input from the Atlantic to the west of Bermuda can be compared with the anthropogenic flux of sulphur calculated to leave the USA (3.9×10^{12} gS yr^{-1})[3]. We estimate the sea area to the west of Bermuda that is bounded to the north by the Canadian border and in the south by the southern tip of Florida, to be $\sim 2.5 \times 10^{12}$ m^2. By scaling down the global ocean estimate[10] of 35×10^{12} gS yr^{-1}, we estimate that the oceanic emission from the sea area to the west of Bermuda could be $\sim 0.2 \times 10^{12}$ gS yr^{-1}. However, this is probably a maximum value as primary production in this area of the Sargasso Sea is much lower, without upwellings such as those found along the west coast of Africa where many of the measurements of Nguyen *et al.*[10] were made. This figure for sulphur produced by natural marine processes is at least 20-fold lower than the anthropogenic sulphur input calculated by Galloway and Whelpdale[3], thus we consider an anthropogenic source for the excess sulphur in Bermuda's rainwater to be the more important. Indeed, aerosol measurements over Bermuda have suggested the presence of anthropogenic sulphate and some trace metals[11,12].

In addition, the major weather patterns in the North Atlantic move in an easterly direction, thus transporting the USA air mass towards Bermuda[9]. The meteorology of the Bermuda area of the North Atlantic Ocean is dominated by the semipermanent sub-tropical Azores/Bermuda high. During the summer and autumn, this high pressure can prevent many of the frontal systems originating over North America from reaching Bermuda. During winter, with a weak or absent sub-tropical high, frontal systems pass regularly. This situation becomes firmly established over the islands during December and persists for about 3 months. Late autumn and spring are the transitional seasons. During June to December, *p*H values are generally high and the excess sulphate, nitrate and ammonia concentrations are relatively low. However, from December to May the reverse is true (Fig. 1). This agrees with the general meteorological pattern described above, although July 1980 and February 1981 represent exceptions. In July an unusually intense storm left the United States coast and reached Bermuda, stalling over the island and producing exceptionally heavy (20 cm over 3 days) and acidic (*p*H 4.4–4.7) rain which biased heavily the monthly average (Fig. 1). In February 1981, a blocking high pressure was established temporarily, bringing unseasonably mild weather for the first 3 weeks of the month. This North Atlantic meteorological effect is seen in the chemistry of the rain for these months.

Figure 2 shows the average *p*H of rainwater associated with air masses arising in each of the three marked sectors as derived

Table 1 Volume weighted mean annual concentrations of major constituents in Bermuda rain from 1 May 1980 to 30 April 1981

Ion	Conc. (μEquiv l^{-1}) in all wet-only samples
H^+	21.15
SO_4^{2-}	43.0
NO_3^-	6.31
Cl^-	221
NH_4^+	4.22
K^+	4.77
Mg^{2+}	46.1
Na^+	178
Ca^{2+}	18.0
Total cations	273
Total anions	270
$^*SO_4^{2-}$	21.54
K^{+*}	0.99
Mg^{2+*}	5.60
Ca^{2+*}	10.15
$^*SO_4^{2-}/NO_3^-$	3.40
NO_3^-/NH_4^+	1.50
Na^+/Cl^-	0.81
$^*SO_4^{2-}/NH_4^+$	5.10

* Non-sea salt components.

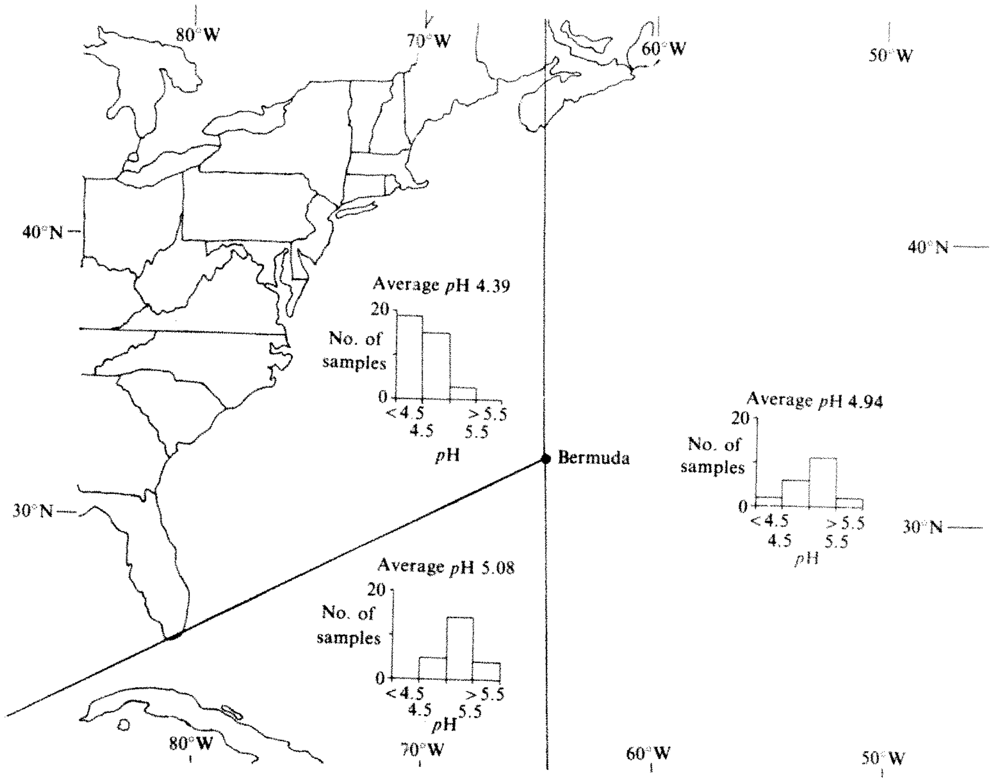


Fig. 2 Map of the north-west Atlantic Ocean showing the three sectors used for ARL back trajectories of Bermuda rain. The average pH of rain originating in each sector is indicated and the distribution of pH values for rain samples from that sector are shown in the histograms.

from ARL back trajectories[12,13]. Clearly, the rainwater originating over continental North America is markedly more acidic than rainfall originating in the other sectors.

A comparison of the composition of rainfall in Bermuda with recently published data from the eastern United States[14] reveals that in addition to a large increase in sea salt concentrations and a reduction in acidity by a factor of 3 during transit from the US coast, there is an increase in the $^{*}SO_4/NO_3$ ratio, indicating an increased relative importance for sulphuric acid. There are several possible reasons for the selective long-range transport of sulphuric as opposed to nitric acid rain precursors to the western North Atlantic from North America—these are discussed elsewhere[15].

In addition to demonstrating the existence and causes of acid precipitation on Bermuda, our results suggest that the dispersion and neutralization of acid rain exported from continents may be a slow process. Over the North Atlantic to the west of Bermuda this appears to be due to the absence of suitable neutralizing agents in the troposphere. In view of this finding, we believe that long-range atmospheric transport of acid rain is possible. Indeed there is growing evidence of acid rain on remote areas[16] although the cause of this acidity is often not as clearly defined as for Bermuda.

We thank Susan Boyd for assistance; Commander M. J. Nemkosky Jr and his staff for meteorological information; and Mr A. Atwood and his staff at Bermuda Harbour Radio. This work was supported by the Bermuda Government, British Petroleum Trading Co. Ltd., ESSO Bermuda, US NOAA, EPA, and US DOE. This is contribution no. 891 from the Bermuda Biological Station and contribution no. 1 to the Global Precipitation Chemistry Programme.

Received 30 November 1981; accepted 26 February 1982.

1. Likens, G. E., Wright, R. F., Galloway, J. N. & Butler, T. J. *Scient. Am.* **241,** 43–51 (1979).
2. Whelpdale, D. M. in *Proc. Regional and Global Observation of atmos. Pollut. Relative to Climate* No. 14, 314–319 (WMO 549, 1980).
3. Galloway, J. N. & Whelpdale, D. M. *Atmos. Envir.* **14,** 409–417 (1980).
4. Bodungen, B. V., Hillier, G. B., Jickells, T. D., Smith, S. R. & Ward, J. *Bermuda Biological Station Spec. Publ.* (in the press).
5. Galloway, J. N. & Likens, G. E. *Wat. Air & Soil Pollut.* **6,** 241 (1976).
6. *USNAS, Environmental Sciences,* 199 (Naval Weather Service Detachment, Ashville, North Carolina, 1974).
7. Galloway, J. N., Cosby, B. J. & Likens, G. E. *Limnol. Oceanogr.* **24,** 1161–1165 (1979).
8. Fay, J. A. & Heywood, J. B. *SAE/DOT Conf. Aircraft and the Environment* Pt 2, 12 (Soc. Automotive Engineers, New York, 1971).
9. Nemkosky, M. J. Jr *Local Area Forecaster's Handbook,* 42 (Naval Oceanographic Command Detachments, Bermuda, 1980).
10. Nguyen, B. C., Gaudry, A., Bonsang, B & Lambert, G. *Abstr. CACGP Symp. on Budget and Cycles of Trace Gases and Aerosols in the Atmosphere,* Boulder, 35 (1979).
11. Meinhert, D. L. & Winchester, J. W. *J. geophys. Res.* **82,** 1778–1782 (1977).
12. Duce, R. A. *et al. Marine Pollutant Transfer* Ch. 3 (eds Windom, H. L. & Duce, R. A.) 77–120 (Lexington, Massachusetts, 1976).
13. Miller, J. M., Knap, A. H., Church, T. M., Galloway, J. N. & Jickells, T. D. in *Symp. on Role of Oceans in atmos. Chem.* (1981).
14. Pack, D. H. *Science* **208,** 1143–1145 (1980).
15. Church, T. M., Galloway, J. N., Jickells, T. D. & Knap, A. H. *J. geophys. Res.* (submitted).
16. Galloway, J. N., Likens, J. E., Miller, J. M. & Keene, W. C. in *Symp. on Role of Oceans in atmos. Chem.* (1981).

Radiations and extinctions of plankton in the late Proterozoic and early Cambrian

Gonzalo Vidal

University of Lund, Micropalaeontological Laboratory, Kemicentrum, Box 740, S-220 07 Lund 7, Sweden

Andrew H. Knoll

Department of Geology, Oberlin College, Oberlin, Ohio 44074, USA

The intrinsic palaeobiological importance of planktonic algal microfossils of Proterozoic and early Cambrian age has been widely documented[1–8]. The evidence supporting the biostratigraphical significance of planktonic microfossils in the correlation of Upper Proterozoic and Lower Cambrian sequences has also been assessed[1,4–8]. We contend here that both within-flora (α) and total (γ) taxonomic diversity trends for late Proterozoic planktonic microfossils indicate that a distinct radiation of presumed cyst-forming eukaryotic plankters occurred during late Riphean and early Vendian (used here in the sense introduced in refs 1, 8). Following this, a mid- to late Vendian extinction episode reduced observable plankton diversity by some 70%, extirpating most morphologically complex, pre-existing taxa. A second radiation restored high plankton diversity levels, but not until well into the early Cambrian.

0028–0836/82/180057—04$01.00

Organic-walled fossils of Proterozoic planktonic microorganisms are preserved in a wide variety of sedimentary rocks, ranging from lagoonal mudstones to distal turbiditic shales and greywackes. Microfossil assemblages can be correlated with sedimentary environments, permitting the palaeoecological distributions of both taxa and diversity patterns to be recognized[2–4]. The stratigraphical ranges of many late Proterozoic plankters are similarly discernible and from these has emerged a distinctive series of informal assemblage zones that makes possible the confident biostratigraphical subdivision and correlation of Upper Proterozoic sedimentary sequences[2,4–8]. Because it is possible to observe both diversity patterns within a single time plane and biostratigraphical distributions in successive time intervals, it should be possible to examine planktonic diversity trends through late Proterozoic time. Such analysis reveals both expected patterns of taxonomic diversification and a hitherto unsuspected decrease in taxonomic diversity that took place near the end of the Proterozoic era.

The planktonic nature of Proterozoic microfossils can be inferred on at least four grounds:

(1) The fossils in question tend to be large (10–>500 μm), robust walled, and, often, ornamented vesicles (Fig. 1). Several Proterozoic taxa further exhibit regular median wall splits, polygonal wall openings, or operculate structures, all of which probably represent excystment mechanisms[2,4,7,8]. In all of these morphological traits, the Proterozoic fossils are closely comparable to Phanerozoic fossil plankton, as well as to the zoosporangia of extant cyst-forming plankters such as prasinophycean green algae[8].

(2) The conditions of occurrence and environmental distribution of Proterozoic plankters are also quite similar to patterns reported from their Phanerozoic counterparts[8]. Palaeozoic[9], Mesozoic[10], and modern phytoplankton species tend to have wide facies distributions, but biotas can be resolved into inshore and offshore associations. The same is true for late Proterozoic plankters[8]. Inshore and, especially, lagoonal biotas are characteristically low diversity assemblages dominated by one or a few taxa, while contemporaneous assemblages from open shelf rocks include a diverse and heterogeneous array of morphologically complex forms[2,3,7,8].

(3) When observable, the distribution of Proterozoic plankters is seen to be independent of the distributions of preserved benthic microbial associations.

(4) As viewed in petrographical thin section, fossils interpreted as planktonic are generally more or less randomly distributed within rock samples, in sharp contrast to known microbial mat builders and other microbenthos which tend to be distributed along bedding.

In general, the morphological and palaeocological criteria outlined above parallel those used to infer the nektonic or planktonic nature of such extinct Phanerozoic groups as ammonites, conodonts, or graptolites. Although there is room for error in interpretation, we are confident that virtually all of the microfossils included in this study do represent planktonic organisms that inhabited coastal or shelf Proterozoic seas.

The taxonomic literature on Proterozoic planktonic microfossils contains numerous synonymies, descriptions of degradational forms, contaminants, and pseudofossils, and other problems that preclude the simple lifting of species lists from published reports. Therefore, for our analysis, we examined sequences of microfossil assemblages spanning the interval from the beginning of the Upper Riphean[1] to the mid-Upper Cambrian, using only those species whose type specimens have been examined by one of us[1] (G.V.). Successions examined come from East Greenland, northern and southern Norway, Sweden, the Russian Platform, the southern Urals, and Svalbard. Taxonomic diversity was compiled at the specific level (genera and suprageneric taxa are decidedly artificial), species being here defined, as in most other palaeontological applications, as the smallest morphologically definable units that are consistently recognizable and differentiable in

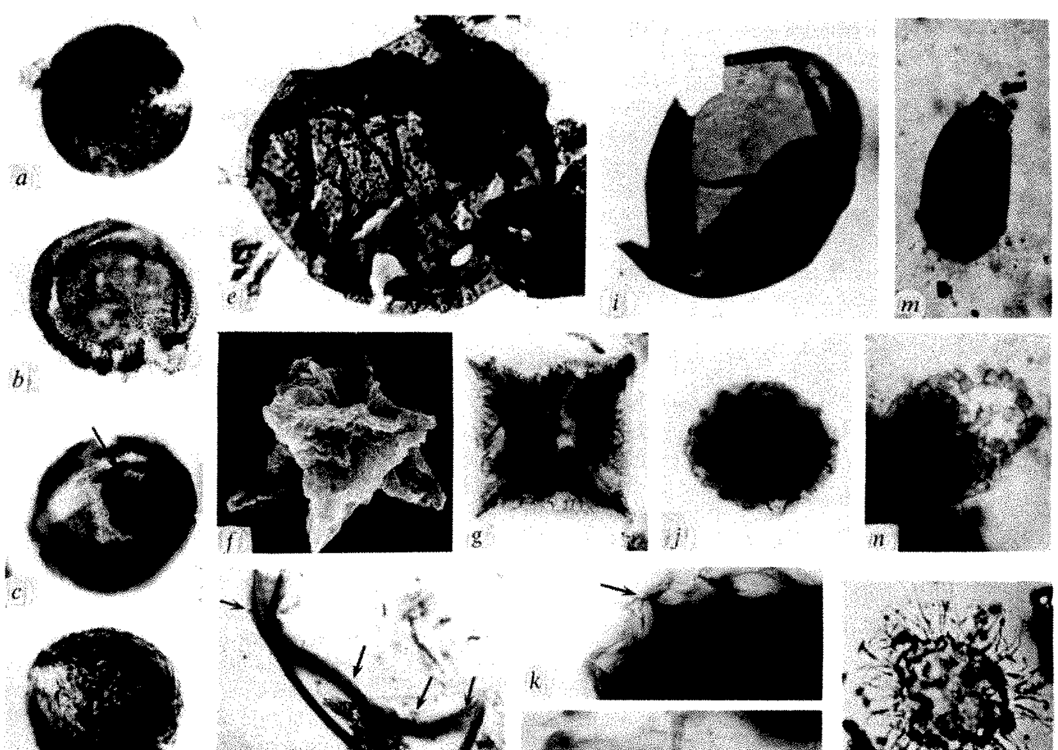


Fig. 1 Representative late Proterozoic and early Cambrian planktonic microfossils. Bar above *o*, 50 μm for *a* and *k*; 25 μm for *b–d*, *g*, *j*, and *m*; 30 μm for *e* and *o*, 20 μm for *f* and *i*, 140 μm for *h*, 70 μm for *l*, 12 μm for *n*. (*a*) *Favososphaeridium favosum* Timofeev, Lower Vendian units of Visingsö Beds, Sweden; (*b*, *c*) *Trachysphaeridium laufeldi* Vidal, Upper Riphean units of Visingsö Beds, Sweden (arrow in *c* points to apparent operculate excystment structure); (*d*) *T. laminaritum* (Timofeev) Vidal, same locality; (*e*) *Kildinella* sp. (*b* of ref. 4), Lower Vendian Ekkerøy Formation, East Finnmark, Norway; (*f*, *g*) SEM and optical views of *Octoedryxium truncatum* (Rudavskaya) Vidal, Lower Vendian units of the Visingsö Beds, Sweden; (*h*) and (*l*) *Trachyhystrichosphaera* sp. (undescribed species) Upper Riphean Hunnberg Formation, Svalbard (arrows in *h* show spines protruding from robust inner wall; *l*, an enlargement of the lower surface of *h*, showing columnar spines and the thin outer membrane they support); (*i*) *K. lophostriata* Jankauskas, Upper Riphean Kwagunt Formation, Arizona (note striate wall structure); (*j*, *k*) *Vandalosphaeridium reticulatum* (Vidal) Vidal, Lower Vendian units of Visingsö Beds, Sweden (*k* is a detail of *j*, showing the internal supports for the outer wall of the fossil); (*m*) vase-shaped heterotrophic protist, Upper Riphean Backlundtoppen Formation, Svalbard; (*n*) *Bavlinella faveolata* (Shepeleva) Vidal, Upper Proterozoic Mineral Fork Formation, Utah; (*o*) *Baltisphaeridium ornatum* Volkova, Lower Cambrian *Mickwitzia* Sandstone, Sweden.

both time and space[11]. The diversity levels plotted for successive stratigraphical intervals in Fig. 2 represent total (γ) diversity; however, almost all forms examined are cosmopolitan and the within-flora (α) diversity trends for each of our several localities closely mirror the figure presented. Note that while our study centred on the peri-North Atlantic region, sequences in part time-equivalent from as far away as Arizona and Namibia exhibit comparable trends. Further, although individual taxonomic decisions may be disputed by other students of early fossils, we believe that the overall shape of the curve, that is, its qualitative characteristics, is accurate and reflects real events in early plankton evolution.

All but one of the microfossil species used in Fig. 2 (*Bavlinella faveolata*) are undoubtedly eukaryotic, and so Fig. 2 can be viewed as a depiction of the early evolution of nucleated cells. The oldest microfossils that can be interpreted with some confidence as probable eukaryotes are found in ~1,400-Myr old shales of the lower Belt Supergroup, Montana[12]. In addition to filamentous cyanobacterial sheaths and thin-walled spheroidal vesicles (possibly envelopes), this microbiota contains what may be the oldest robust-walled acritarchs. Data for the ensuing Middle Riphean interval are sparce, but available information suggests that diversity rose gradually during this period.

As is evident from Fig. 2, Upper Riphean sequences are characterized by the widespread expansion of sphaeromorphic acritarchs displaying a wide variety of sculptural patterns and possibly excystment mechanisms[7,8]. The succeeding Lower Vendian[1] (perhaps equivalent to the uppermost Riphean Kudashian of the southern Urals) biota is even more diverse. The significance of this expansion is perhaps best understood by analogy to the well known Cretaceous diversification of angiosperm pollen taxa documented by Doyle[13]. Arguments over flowering plant antiquity long centred on the interpretation of scattered and, often, poorly preserved individual plant fossils. Doyle demonstrated that no matter when angiosperms first evolved, the early Cretaceous period marked the evolutionary radiation and ecological rise to dominance of the group. We contend that the late Riphean diversification of marine acritarchs tells a similar story. It is unlikely that this radiation documents the initial appearance of eukaryotic cellular organization. Indeed, most models for eukaryogenesis call for early cells that were wall-less and, hence, virtually unfossilizable. The late Proterozoic expansion does, however, record the initial expansion and rise to ecological importance of presumably cyst-forming eukaryotic plankton, and, thus, the initial radiation of nucleated organisms observable in the fossil record. As far as known, eukaryotes radiated in the late Riphean and early Vendian intervals and, insofar as the fossil record permits us to compare different environments, first established dominance in the planktonic realm[15]. By the end of the Riphean era, the preservable plankton biota included vase-shaped heterotrophic protists, as well as algae[16,17].

A drastic change in the marine plankton biota occurred in mid-Vendian (Varangerian[18]) times, coincident with the widespread Varangerian glacial epoch[8]. Almost 70% of pre-existing taxa disappeared, including almost all of the morphologically complex early Vendian acritarchs and leaving a depauperate biota of simple sphaeromorphs and multi-sphere microfossils interpreted as endosporulating cyanobacteria (*B. faveolata* (Shepeleva) Vidal[2]). *B. faveolata* is known to have dominated glacially influenced marine environments, probably proliferating in ecologically stressed conditions[18,19].

Although the Varangerian plankton record leaves much to be desired because of associated erosional contacts and hiatuses, the reality of the extinctions can be supported on several grounds. (1) Marine Varangerian rocks do exist and contain low diversity assemblages. Such within flora diversity does not depend on outcrop representation in the same way as does total diversity (at least for biotas in which not all forms are cosmopolitan)[20]. (2) One can argue that observed low Varangerian plankton diversities result from a biased sampling of environments—that we are looking at glacially influenced sequences. This is true for the most part; however, immediately superjacent Valdaian (Upper Vendian) strata are equally species poor, and these deposits are widespread and well sampled. Taxa that disappeared at the end of the early Vendian (=Kudashian?) period simply did not reappear, even though non-glacially influenced environments appropriate for their growth and preservation returned. (3) The taxa that disappeared apparently became extinct without issue. Not only do diagnostic acritarchs like *Trachysphaeridium laufeldi* and *Kildinella lophostriata* (Fig. 1) disappear at the end of the upper Riphean–early Vendian period, but younger acritarch assemblages contain no fossils that seem to belong to the same phylogenetic clade. Thus, there is little doubt that the latest Proterozoic eon witnessed a significant extinction episode in the microplanktonic realm. It is tempting to relate this biological event to the Varangerian glaciation(s) because of the time coincidence; however, any hypothesis relating the two phenomena must be speculative.

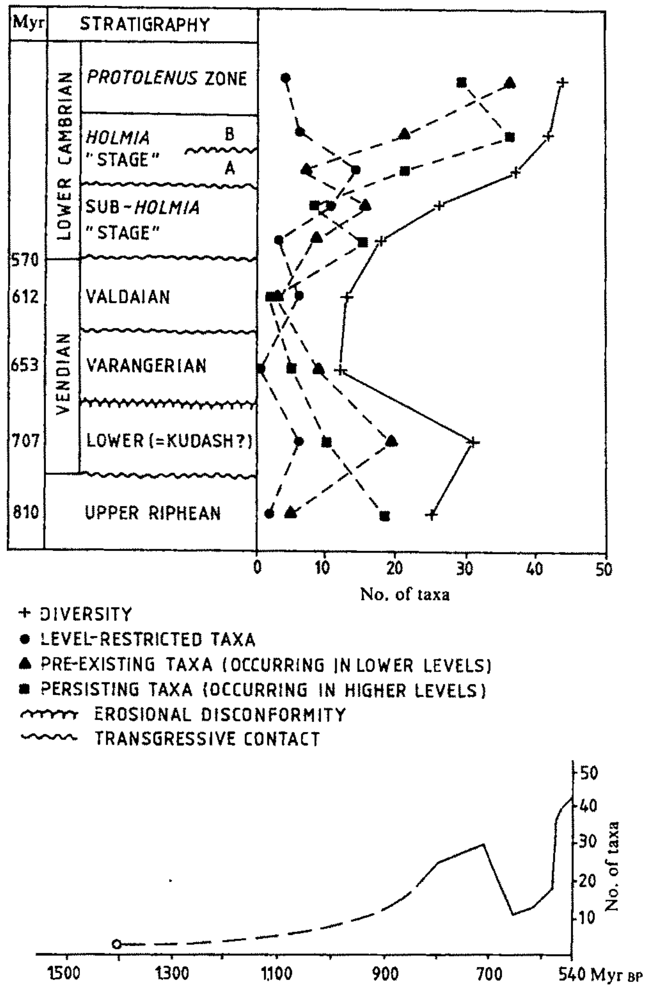


Fig. 2 Diversity changes in the late Proterozoic and early Cambrian plankton record. Upper chart shows details of diversity changes near the Proterozoic–Cambrian (wavy lines between stratigraphical intervals indicate transgressive contacts in Scandinavia; saw-tooth lines indicate erosional disconformities). The lower chart illustrates diversity changes within a broader framework of time. The dotted line is an estimate of the course of eukaryotic plankton diversity from 1,400 Myr BP, the date marking the oldest known probable eukaryotic plankters, to the Upper Riphean (where data become quantitatively reliable). The charts originate from published and unpublished data of continuous Upper Proterozoic–Lower Cambrian sequences in Greenland, Scandinavia (including Svalbard), the Russian Platform, the Urals, and so on, comprising thousands of investigated samples. The microfossil taxa on which the numerical data are based are listed in refs 1, 2, 4–6, 17, 18, 19, 21–23. The Vendian is here used in the sense proposed in refs 1, 8, and comprises pre-Varangerian, Lower Vendian, units in Scandinavia which may perhaps be equivalent with the terminal Riphean (Kudashian) of the southern Urals in the USSR. Definitive proofs for this are, however, missing.

Plankton diversity rebounded sharply in early Cambrian times (Fig. 2), with the radiation of spinose forms attributed to the Palaeozoic acritarch genus *Baltisphaeridium* and other new and morphologically complex acritarch forms. When it occurred, the Cambrian radiation was remarkably rapid; however, note that the most comprehensive radiation does not coincide in time with the beginning of the Cambrian period[21]. Plankton diversity underwent a slow increase throughout the Sub-*Holmia* 'stage'[22,23] of the lower Cambrian system.

We conclude that the fossil record of early eukaryotic plankton has a distinctive and not entirely predictable shape. Late Riphean and early Vendian microplankton assemblages indicate a dramatic morphological (taxonomic) diversification that represents the first recognizable radiation of eukaryotic algae and documents their rapid rise to dominance in coastal planktonic habitats. An equally striking drop in diversity near the end of the Proterozoic era is interpreted as the first recognizable episode of widespread extinction. More data, particularly from sedimentary sequences outside the peri-North Atlantic area will be required to test and refine the arguments presented here.

This work was supported by grants from the Swedish Natural Science Research Council (N.F.R.) and the Knut och Alice Wallenberg Stiftelsen (G.V.) and NSF grant DEB 80-04290 (A.H.K.).

Received 7 January; accepted 17 March 1982.

1. Vidal, G. *Publ. Inst. Mineral. Palaeont. Quat. Geol. Univ. Lund* **219**, 1–22 (1979).
2. Vidal, G. *Fossils Strata* **9**, 1–57 (1976).
3. Knoll, A. H. in *Paleobotany, Paleoecology, and Evolution* Vol. 1 (ed. Niklas, K. J.) 17–54 (Praeger, New York, 1981).
4. Vidal, G. *Norges Geol. Unders.* **362**, 1–53 (1981).
5. Timofeev, B. V. *Mem. 129*, 1–350 (V.N.I.G.R.I., Leningrad, 1959); *Acad. Sci. U.S.S.R., Lab. Precambrian Geol.*, 1–145 (Nauka, Leningrad, 1966).
6. Timofeev, B. V., Hermann, T. N. & Mikhajlova, N. S. *Acad. Sci. U.S.S.R., Inst. Geol. Geochronol. Precambrian*, 1–106 (Nauka, Leningrad, 1976).
7. Vidal, G. & Knoll, A. H. *Geol. Soc. Am. Mem.* (in the press).
8. Vidal, G. *Precamb. Res.* **15**, 9–23 (1981).
9. Jacobson, S. R. *J. Paleontol.* **53**, 1197–1212 (1979).
10. Wall, D. *Micropaleontology* **11**, 151–190 (1965).
11. Niklas, K. J., Tiffney, B. H. & Knoll, A. H. *Evolut. Biol.* **12**, 1 (1980).
12. Horodyski, R. J. *J. Paleontol.* **54**, 649–663 (1980).
13. Doyle, J. A. *J. Arnold Arbor.* **50**, 1–35 (1969).
14. Doyle, J. A. & Hickey, L. J. in *Origin and Early Evolution of Angiosperms* (ed. Beck, C. B.) 139–206 (Columbia, New York, 1976).
15. Knoll, A. H., in *Biotic Interactions in Recent and Fossil Benthic Communities* (eds Tevesz, M. J. & McCall, P. L.), (Plenum, New York, 1982).
16. Bloeser, B., Schopf, J. W., Horodyski, R. J. & Breed, W. J., *Science* **195**, 676–679 (1977).
17. Knoll, A. H. & Vidal, G. *Geol. För. Stockh. Förh.* **102**, 207–211 (1980).
18. Vidal, G. *Grønl. Geol. Unders. Rapp.* **78**, 1–19 (1976).
19. Knoll, A. H., Blick, N. & Awramik, S. M. *Am. J. Sci.* **218**, 247–263 (1980).
20. Bambach, R. K. *Paleobiology* **3**, 152–167 (1977).
21. Volkova, M. B. *et al. Acad. Sci. U.S.S.R. Geol. Inst.*, 1–210, (Nauka, Moscow, 1979).
22. Vidal, G. *Geol. Fören. Stockholm Förh.* **103**, 183–192 (1981).
23. Pashchkyavichene, L. T. *Acad. Sci. U.S.S.R. Paleontol. Inst.*, 1–74 (Nauka, Moscow, 1980).

Mating and song types in the great tit

P. K. McGregor & J. R. Krebs

Edward Grey Institute, Department of Zoology, South Parks Road, Oxford OX1 3PS, UK

We report here the first field investigation of the relationship between the songs experienced by female songbirds early in life and those produced by their mates. Our 6-yr study of the great tit (*Parus major*) showed that females in a ringed population tended to be mated with males singing slightly 'unfamiliar' songs, where 'familiar' songs are defined as those sung by the female's father. This result parallels recent findings on sexual imprinting in birds[1,2].

We recorded the songs of all the breeding male great tits (between 18 and 33 birds in a year) in Marley Wood, Oxford, during 1975–80. Most of the breeding adults and all their young were individually ringed throughout the study. The recordings were made before and during the breeding season (March–May) by visiting all the territorial males several times to record both spontaneous song and song produced in response to playback tapes. Each male had a repertoire of one to six song types ($\bar{x} = 2.98$, $n = 152$). We classified the 454 songs recorded into 41 categories by inspecting narrow band sonagrams. Great tit song is a loud vocalization consisting of rhythmically repeated units (phrases) which are uttered in bursts with intervals of silence between. The major criteria for classifying songs were as follows[3]: (1) number of notes per phrase, which varies over 1–11; (2) type of note (pure or modulated); (3) frequency relationship of notes (for example, in a song with two note types, one pure, one modulated, the pure note could be of higher or lower frequency than the modulated; these two possibilities would count as two different song types); (4) duration of song phrase. We placed less emphasis on variation in duration of notes and modulation rate within notes. When P.K.M. and J.R.K. classified songs independently they obtained 85–95% agreement. The song types about which we did not agree were usually rare ones sung by only one bird in a year and we assigned these to new categories.

Our analysis is based on 17 females for which we knew the repertoires of both husband and father. We compared the songs of the 17 'pairs' of female's fathers and mates in two ways: first by looking at the number of song types shared, and second by

Table 1 Mating and song sharing in the great tit

No. of songs shared with female's father

a

	0	1	>1
Husbands	6	11	0
Other males	242	157	45

b

	0	1	>1
Husbands	6	11	0
Other males	84	50	10

c Degree of difference between father and husband*

	'Similar'	'Intermediate'	'Different'
Observed (husbands)	3	5	1
Expected (other males)	3·15	1·8	4·05

a, The number of females mated to males sharing 0, 1 and more than 1 song with the female's father. This distribution was compared with one obtained by counting all the males (except for fathers) in Marley Wood in the year in which the female in question bred. The table shows overall heterogeneity ($\chi^2_{2\,d.f.} = 6.8$, $P<0.05$), and a comparison of 'one song shared' versus 'no songs' or 'more than one song' is significantly different from expected ($\chi^2_{1\,d.f.} = 6.09$, $P<0.02$). *b*, A similar analysis to that in *a*, but considering only males that nested between 201 and 400 m from the female's birth site. The table shows overall heterogeneity ($\chi^2_{2\,d.f.} = 6.6$, $P<0.05$). Comparison of 'one song shared' versus 'no songs or more than one' $\chi^2_{1\,d.f.} = 5.82$, $P<0.02$. *c*, Of the 17 husbands analysed above, 11 shared one song with the female's father. In nine of these cases other males in the husband's year also shared the song with the father. For these nine songs it was possible to compare the difference in detailed song structure between father and husband with that between father and other males. Husbands fell into the category 'intermediate difference' slightly more often than expected ($P = 0.06$, Fisher exact test). It is reasonable to test the table directly for an excess of 'intermediate' mates because *b* predicts this.

* (1) Take six measures of song structure for all males (including husband) that share the song type in question with female's father. (2) Rank males in order of similarity to father on each measure. (3) Divide rank orders in three equal-sized categories: 'similar', 'intermediate', 'different'. (4) For each male determine which category is commonest. (Ties: share between tied categories.) Use this as estimate of overall similarity. (5) Compare observed distribution of similarity between husbands and fathers with that expected based on all other males.

examining the detailed structure of any songs that were sung by both males.

Table 1*a* shows that females are more likely than expected by chance to be mated with males that share one song type with the father. There is also a slight but nonsignificant suggestion that the 17 females avoid mating with males possessing 'rare' songs (those used by four or fewer males in a year). The observed number of females mated to a male with more than one rare song was 1 out of 17, while the expected number was 3.55 out of 17 ($P \sim 0.1$ Fisher exact test). Rare songs tend to be sung by immigrant males from outside the wood[3]. The females were all born within the wood, hence the observed trend suggests avoidance of mating with males from other populations. This conclusion might be spurious if we have overestimated the availability of unpaired immigrant birds, as might happen, for example, if immigrant birds (mostly young ones) arrive already paired. However, young birds disperse in July–August[4] and usually do not pair until September–February[5], which suggests that any non-randomness of pairing arises after, not before, birds immigrate.

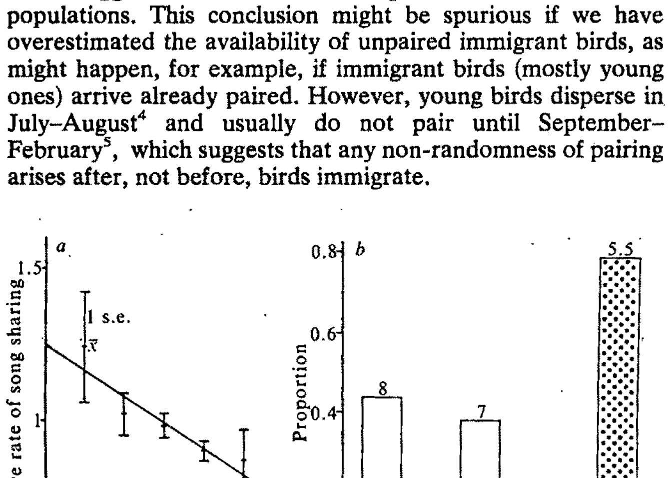


Fig. 1 *a*, The correlation between the rate of song sharing [(N/R), where N = number of songs shared and R = repertoire size of the reference male] by males and the distance between their territories. There are five distance categories and we measured the rate of song sharing in each category for 6 yr (therefore $n = 5 \times 6 = 30$). The results are expressed as rates of song sharing relative to the mean for each year in order to eliminate between-year differences in mean rate. The mean number of songs shared between neighbours (distance category 1) is 0.65 songs. *b*, The similarity of songs shared by fathers and males within the Marley population (open histograms), and fathers and males outside Marley Wood (shaded). The height of the histograms indicates the number of males in each similarity category (see text for calculation) as a proportion of all Marley males ($n = 18$) or all non-Marley males ($n = 7$). Actual numbers are given above each histogram, that is, eight Marley males share a song type which is the same as the father's and is similar in detailed structure. Only two song types were found within and outside Marley; the results for these have been combined.

Table 1 computes the expected distribution by counting the number of males sharing 0, 1 and >1 songs with the father. In this calculation we pooled all the males in Marley Wood during the year in which the female in question bred. This assumes as a null hypothesis that a female is equally likely to mate with any male in the wood. However, it is known that females undertake dispersal from their natal site before breeding and are hence unlikely to mate with males nesting very close to the father's territory. Furthermore, the rate of song sharing between males is correlated with distance between their territories (Fig. 1*a*). We therefore refined our null hypothesis by calculating the expected distribution using only males living between 201 and 400 m from the females' birth site. This corresponds to the average dispersal distance of three territories for females calculated by Greenwood *et al.*[6]. Table 1*b* shows that females are less likely than expected to mate with males sharing 0 or >1 song with the father. However, the slight tendency to avoid husbands with rare songs disappears when dispersal is taken into account (the revised expected number is 0 out of 17, the observed is out of 17).

By correcting for dispersal distance, we have tried to eliminate the possibility that our results could be explained solely as a by-product of dispersal distance. This does not, however, show that song and dispersal are unrelated: for example, young birds could be sensitive to the changing frequency of song types with distance (Fig. 1*a*) in choosing where to settle.

Table 1*b* can be considered in two parts in order to assess the contribution of avoidance of similar and dissimilar repertoires to the final result. Comparison of columns 2 and 3 of Table 1*b* shows only a slight suggestion of avoidance of males with very similar repertoires to the father ($\chi^2_{1 \text{d.f.}} = 2.13$) and a similar comparison of columns 1 and 2 shows that significantly fewer than expected husbands occur in column 1 ($P < 0.05, \chi^2_{1 \text{d.f.}} = 4.7$). Therefore, both effects contribute to the overall level of significance, but avoidance of dissimilar repertoires has a greater effect.

Table 1*c* shows that the bias in favour of 'intermediate degree of difference' shown above for song types also applies to detailed song structure. For 9 of the 11 husbands which shared a song type with the female's father we could compare the degree of difference between husband and father with that between fathers and other males in the population (see Table 1 legend for details of the method). Slightly more females than expected were mated to males of intermediate difference. As in Table 1*b*, we corrected for dispersal distance in constructing our null hypothesis, as the degree of similarity of detailed structure varies with distance between territories[17] (Fig. 1*b*).

Thus, our results are best explained by the idea that females prefer a balance between similarity and dissimilarity. The trends we have observed could be summarized as 'prefer males that share one song type with father, prefer shared song to be slightly different from father's'. A parallel finding for plumage patterns has been found in experiments on sexual imprinting in quail[1,2,7]. This has been interpreted as indicating that females might strike an optimal balance between inbreeding and outbreeding, an effect described in plants[8]. Avoidance of inbreeding[9,10] (which is known to be disadvantageous in great tits[11]) and outbreeding[12] has previously been suggested as a possible consequence of song learning in birds. Although our results are consistent with the predictions of optimal outbreeding, we must still test the three main assumptions of the hypothesis: (1) Females learn paternal songs. (2) Song types influence mate choice in the great tit, as found in other species[13,14]. (Song seems to be important in mate attraction in the great tit because males whose mates have been removed sing more than paired males at the same time of year[15].) (3) There is a cost of outbreeding between populations, as has been shown in some other species[16,18].

We acknowledge the SRC for financial support and thank Pat Bateson, Peter Slater, Paul Sherman, Paul Harvey and Chris Perrins for their comments on the manuscript and Andy Dobson, Alan Grafen and Michael Bulmer for statistical advice.

Received 19 November 1981; accepted 23 March 1982.

1. Bateson, P. P. G. *Nature* **273**, 659–660 (1978).
2. Bateson, P. P. G. *Z. Tierpsychol.* **53**, 231–244 (1980).
3. McGregor, P. K. & Krebs, J. R. *Behaviour* (in the press).
4. Goodbody, I. M. *Br. Birds* **45**, 279–285 (1952).
5. Hinde, R. A. *Behav. Suppl.* **2**, 1–201 (1952).
6. Greenwood, P. J., Harvey, P. H. & Perrins, C. M. *J. Anim. Ecol.* **48**, 123–142 (1979).
7. Bateson, P. P. G. *Nature* **295**, 236–237 (1982).
8. Price, M. V. & Waser, N. M. *Nature* **277**, 659–660 (1978).
9. Pulliam, H. R. & Dunford, C. *Programmed to Learn*, 50–52 (Columbia University Press, 1980).
10. Lumsden, C. J. & Wilson, E. O. *Genes, Mind and Culture. The Co-evolutionary Process* (Harvard University Press, 1981).
11. Greenwood, P. J., Harvey, P. H. & Perrins C. M. *Nature* **271**, 52–54 (1978).
12. Baker, M. C. *Evolution* **29**, 226–241 (1975).
13. Searcy, W. A. & Marler, P. *Science* **213**, 926–928 (1981).
14. Baker, M. C., Spitler-Nabors, K. J. & Bradley, D. C. *Science* **214**, 819–821 (1981).
15. Krebs, J. R., Avery, M. I. & Cowie, R. *Anim. Behav.* **29**, 635–637 (1981).
16. Shields, W. M. in *The Ecology of Animal Movement* (eds Swingland I. R. & Greenwood, P. J.) (Oxford University Press, 1981).
17. Falls, J. B., Krebs, J. R. & McGregor, P. K. *Anim. Behav.* **30** (in the press).
18. Levin, D. A. *Ann. Miss. bot. Gard.* **68**, 233–253 (1981).

Male *Eufriesia purpurata*, a DDT-collecting euglossine bee in Brazil

Donald R. Roberts*†, Wilson D. Alecrim*‡, Jack M. Heller§, Susan R. Ehrhardt§ & Jose B. Lima*

* Nucleo de Medicina Tropical e Nutricao, Faculdade de Ciencas da Saude, Universidade de Brasilia, Brasilia, Brasil

§ Pest Management and Pesticide Monitoring Division, US Army Environmental Hygiene Agency, Aberdeen Proving Ground, Maryland 21010, USA

While studying the ecology of the malaria vector *Anopheles* (*Nyssorhynchus*) *darlingi* Root along the Ituxi River, Amazonas, Brazil, we observed aggregates of bees on the walls of houses that were routinely sprayed with DDT. Several bees collected from DDT-treated house walls in August 1978 were identified as male specimens of *Eufriesia purpurata* (Moscary) of the tribe Euglossini (Hymenoptera: Apoidae). (The bees were identified as *Euplusia purpurata* by Anthony Raw, Laboratorio de Ecologia, Universidade de Brasilia. The change of generic name from *Euplusia* to *Eufriesia* is based on ref. 1.) These bees were well known to the local residents as the insects that eat DDT and we present here the first documentation that they (1) are attracted to DDT, (2) actively collect large quantities of DDT from treated house walls and (3) suffer no apparent insecticidal effects. We also found that the frequency of house visiting is most intense during July to September. Most bees arrive at houses before 12.00 h, remain 2–3 h and return on subsequent days to collect more DDT. Noise produced by bees as they collected DDT was a notable disturbance to 76% of 21 families interviewed along the Ituxi River.

The euglossine bees have been extensively studied due to their important role as plant pollinators. Both males and females provide valuable pollination services, but a more specific relationship has been found for the males in their pursuit of certain flower odours, and many flowers only attract a single bee species that affects pollination[2,3]. The attractants of male euglossine bees are characteristically aromatics and monoterpenes[3].

Five bees collected in 1978 were analysed for DDT residues in accordance with analytical procedures recommended by the United States Environmental Protection Agency[4]. The head, thorax, abdomen and fore-, mid- and hindlegs were processed separately for each specimen (wings were discarded). The body parts were weighed, before DDT extraction, to obtain concentrations of insecticide in parts per million (p.p.m. or μg per g) by body weight.

The sum of the DDT isomers expressed as DDTR, converted to p.p.m. by dry weight, revealed exceptionally high concentrations of DDT (Table 1). Highest concentrations were consistently found in extracts of the hindlegs, indicating that DDT was stored in the hind tibial pouch. This finding was consistent with observations on the storage site of other attractants collected by euglossine bees[3].

During investigations made from September to October 1979, the attractancy of DDT to males of *E. purpurata* was demonstrated by exposing two treated boards for each of the major components of the standard DDT wettable powder formulation. The standard DDT powder consisted of 75% DDT, talc and a wetting agent and was sprayed on house walls as a suspension in water. A common house construction material, slats from açai palm, was used for the 22.5 × 9.0 cm test boards. The surface of two boards was wetted for each of the following treatments; (1) three times with 625 p.p.m. DDT (technical grade) in methanol solution; (2) suspension of talc in water; (3) water; and (4) methanol. The specific wetting agent could not be identified and was not included in the test; as it was a minor component of the spray formulation, we decided that the compound should be identified and tested only if there was no apparent DDT attractancy. The treatments described above provided eight boards which were exposed simultaneously to populations of *E. purpurata*.

Treatments were exposed in random block design (two rows of four boards each) on a tree 0.5 m above the ground. Bees were common around the tree because an ant nest at the tree base had been sprayed with DDT in April 1979. Treated boards were randomly assigned to eight fixed positions and were relocated seven times. Visual counts of bees on the boards were made on 3 days. Eighty-three *E. purpurata* landed on the DDT-treated boards, as determined by 23 separate counts, and no bees landed on any other boards. Analyses for DDT surface residues on the treated and untreated boards revealed DDTR concentrations of 1,236 and <1.0 p.p.m., respectively. That the concentrations on treated boards were lower than expected was probably due to leaching of some DDT crystals by a light rain on the last day of observations. These test results demonstrated the specific attractancy of DDT for *E. purpurata* males.

The daily house-frequenting activities of male *E. purpurata* were also studied, either by conducting hourly counts of bees by location in the house or by collecting all bees as they arrived at the house. Bees were collected using aerial nets and numbers captured were recorded at 15-min intervals. Count and collection data were obtained for a 5-day series with hourly counts performed on days 1 and 3. Continuous collections were conducted on days 2, 4 and 5.

The hourly counts produced a mean of 397 bees per day compared with a mean of 135 bees per day captured during 3 days of continuous collections. The collection data indicated that most of the bees that visited the house on a particular day first arrived before 12.00 h. Studies of in-house distributions of bees and insecticide residues revealed that ~78.6% of the bees were located on the inner walls of the house during the 2 days of hourly counts, the remaining 21.4% being observed on the thatch ceiling, exterior walls and wood superstructure of the house. Wall and ceiling samples were collected from four houses to document the within-house distribution of DDT residues. Average DDTR concentrations in p.p.m. (range limits in parentheses) were as follows: upper ceiling, 152 (36–413); lower ceiling, 5,892 (10–9,225); interior wall, 102,101 (40, 670–170,670); exterior wall, 2,120 (35–8,110). Clearly, the within-house site of highest DDTR concentrations corresponded with the site of most intense bee activity.

Twenty-eight bees from these studies were analysed for DDTR concentrations. The geometric mean DDTR concentrations in bees collected immediately on arrival at the house, specimens marked while collecting DDT and captured 24 h later, and specimens marked in the morning and captured in the afternoon, were 1,197, 4,809 and 2,152 p.p.m., respectively. The DDTR concentrations found in bees captured early in the morning, before they could alight on the DDT-treated walls, indicated that many specimens had collected DDT on a preceding day and survived through the night. Furthermore, 3 of about 50 bees that had been marked with an acrylic paint on day 3 were recaptured on day 4.

In July–August 1980, studies were conducted with mark–capture and bioassay methods to define further the effects of

† Present address: Department of Entomology, Walter Reed Army Institute of Research, Washington, DC 20012, USA.

‡ Present address: Instituto de Medicina Tropical de Manaus and Universidade do Amazonas, Estrada Pedro Texeira S/N, Manaus, Amazonas, Brazil.

0028–0836/82/180062—02$01.00

DDT on bee longevity. The bees were individually marked with USR fluorescent pigment 1953 as they collected DDT from the house walls and were identified with a Blak-Ray, ULV:56, long-wave UV lamp on subsequent capture. This procedure was used in a 6-day series of observations to confirm previous findings on the diurnal activity patterns. The bees were counted hourly and 103 specimens were marked on day 1. Continuous collections of bees were made on day 2; this was followed by a repeat of the first day activities on day 3, when an additional 108 bees were marked. Continuous collections were again made on days 4, 5 and 6. Seventy (33%) marked bees were captured, with 20, 19, 16 and 15 being captured on days 2, 4, 5 and 6, respectively. These results clearly indicate that the bees survived the exposure to DDT and returned on subsequent days to collect DDT from sprayed houses.

Results from this 6-day series of observations on the diurnal activity cycles were in general agreement with the 1979 studies. A daily mean of 188.5 bees were recorded during 2 days of hourly counts whereas the mean number captured per day was 89. The main difference was that proportionately more euglossine bees were captured in the afternoon during the 1980 studies; for example, 60% of the total was captured before 12.00 h in 1980 compared with 92% in 1979. Count and collection data were also used to calculate crude estimates of time that the average bee remained in the house by dividing the mean number collected in the house per day by the mean daily number reported in hourly counts—2.9 and 2.1 h for studies in 1979 and 1980, respectively. These estimates clearly indicated that individual bees actively collected DDT for a considerable length of time.

Table 1 Residue levels of DDTR in male *Eufriesia purpurata* collected from sprayed houses along the Ituxi River, Amazonas, Brazil

	Residue levels of DDTR (p.p.m. or μg per g)						
Bee no.	Fore-legs	Mid-legs	Hind-legs	Head	Thorax	Abdo-men	Total p.p.m. (μg per g) per bee
1	452	930	10,526	124	30	119	12,181
2	1,267	4,245	29,083	359	60	203	35,227
3	2,465	6,222	11,203	839	154	121	21,004
4	2,769	1,685	10,631		99		>15,184
5	10,998	3,706	25,026	1,326	399	173	41,627

The bioassay test consisted of comparing mortalities in bee populations that had been exposed to DDT with unexposed populations. In 1979 we discovered that *E. purpurata* males could be attracted to DDT-treated sheets in the forest. Therefore, at a site ~5 km from human habitations, two screen-encased DDT-treated sheets were used to attract bees, which were captured and subsequently used as normal controls in a bioassay test. Twenty-five bees captured from house walls and 18 captured in the forest were placed individually in tubes, provided with cotton pads soaked with 10% sucrose solution and observed for mortality. All bees were held in a Styrofoam ice chest to maintain a high humidity and protect against temperature extremes. After 54 h, 22% of controls had died compared with only 8% mortality among specimens collected from the DDT-treated walls. When this test was repeated in October 1980, the 48-h mortalities were 20 and 28% for the control and test populations, respectively. Our bioassay test results, plus the mark–capture data, gave no indication that the *E. purpurata* males were in any way harmed by high concentrations of DDT. In addition, we have made observations for many days on hundreds of bees as they collected DDT and no specimens demonstrated insecticidal effects (for example, loss of orienta-demonstrated insecticidal effects (for example, loss of orientation, weakness, paroxysms). This is surprising because the average LD_{50} of *p,p*′-DDT for honeybees is 6 μg per bee[5]. In contrast, the average concentration for this isomer in five specimens of *E. purpurata* was 2,039 μg per bee (Table 1).

Head of household interviews were conducted in July 1980 for 21 families along the Ituxi River. All individuals knew of the bees and 71% had bees in their houses at the time of the interviews. In addition, 100% responded that there had been no house-visiting bees until the malaria control spray programme was begun. There was unanimous agreement on the time of year that the bees appeared in significant abundance—July–September. Noise produced by bee activity reportedly was not disturbing to 5% of those interviewed, more or less disturbing to 19% and was a notable disturbance to 76% of the families interviewed. Fortunately, male euglossine bees are stingless.

Euglossine bees have been reported to frequent houses in other areas of the Amazon Basin, and have also been observed brushing on house walls near Iquitos, Peru (R. L. Dressler, personal communication). That this is a widespread phenomenon was verified by a recent report on sightings of *E. purpurata* in houses in many areas in the state of Pará, Brazil[6].

Simple summary statistics suggest that populations of *E. purpurata* remove significant amounts of DDT from sprayed houses. An average of 112 bees per day were collected from a DDT-treated house during studies in 1979–80. If each bee removed 2,459 μg of DDT (value = average dry weight of bee × average DDTR in μg per g = $0.0982 \times 25{,}044.6$), then a total of 24.8 g of DDT, theoretically, could be removed during 3 months of intense bee activity. This would be equivalent to stripping the DDT from 12 m^2 of wall surface. However, as individual bees make repeated visits to houses, we conclude that our estimate for numbers of bees per day is an overestimate and, thus, they probably do not remove a sufficient amount of DDT to have an adverse impact on the malaria control effort.

Observations reported here demonstrate that male *E. purpurata* are attracted by the odour of DDT, which stimulates their behaviour of brushing the insecticide from walls for storage in a pouch of the hind tibia. The reasons for this behaviour are unknown but it may be that DDT or some aspect of its structural conformation is mistaken by the male bees for an odour substance which they normally collect in nature. Such substances could serve as territorial markers or they may be metabolically altered and secreted as sex attractants[3,7,8]. More recent research on the fate of odour substances provides additional evidence that collected substances are metabolized (or altered) by the male bees (N. H. Williams, personal communication). Clearly, many aspects of euglossine biology remain to be clarified.

We thank Professor Aluizio R. Prata, Myron G. Radke and Willis A. Reid Jr for their help and encouragement, Dr Ronald Ward for reading the manuscript and Wilma Rich for typing. This work was supported by the Brazilian Conselho Nacional de Pesquisas and research contract DAMD 17-79-G-9450 from the US Army Medical Research and Development Command, Office of the Surgeon General, Ft Detrick, Frederick, Maryland 21701. The opinions contained herein are those of the authors and should not be construed as official or reflecting the views of the Department of the Army.

Received 20 November 1981; accepted 16 March 1982.

1. Kimsey, L. S. *Pan-Pacif. Ent.* **55**, 193 (1979).
2. Dressler, R. L. *Evolution* **22**, 202 (1968).
3. Dodson, C. H., Dressler, R. L., Hills, H. C., Adams, R. M. & Williams, N. H. *Science* **164**, 1243–1249 (1969).
4. *Manual of Analytical Methods for the Analysis of Pesticides in Residues in Human and Environmental Samples*, Sect. 5, 2 (2) (USEPA, Research Triangle Park, North Carolina, 1974).
5. Atkins, E. L., Greywood, E. A. & Macdonald, R. L. *Toxicity of Pesticides and Other Agricultural Chemicals to Honey Bees* (University of California Agricultural Extension M-16, 1973).
6. Fraiha, H. *Saude na Amazonia* (ANEPS, Sao Paulo, 1979).
7. Kimsey, L. S. *Anim. Behav.* **28**, 996–1004 (1980).
8. Landim, C. D. C., Stort, A. C., Cruz, M. A. D. C. & Kitajima, E. W. *Revta bras. Biol.* **25**, 323–342 (1965).

Erythrocytes deficient in glycophorin resist invasion by the malarial parasite *Plasmodium falciparum*

G. Pasvol, J. S. Wainscoat & D. J. Weatherall

Tropical Medicine Unit, Nuffield Department of Clinical Medicine,
John Radcliffe Hospital, Headington,
Oxford OX3 9DU, UK

It has been suggested that one of the main factors which determines host susceptibility to different malarial parasites is the interaction of their invasive forms, merozoites, with specific receptors on the red cell membrane[1–3]. Thus the Duffy blood group determinants may be involved in the entry of *Plasmodium vivax* but not of *Plasmodium falciparum* into human red cells. When analysing red cells deficient in various blood group antigens, Miller *et al.* noted[3] that two samples of En(a−) cells showed a reduction of invasion by *P. falciparum*[4]. These cells lack both the major transmembrane sialoglycoprotein, glycophorin A (or MN glycoprotein), and the independently segregating Wright[b] (Wr[b]) antigen and also show increased glycosylation of band 3, the major membrane-penetrating glycoprotein. Despite this, En(a−) individuals are haematologically normal[5]. We have now confirmed that En(a−) cells obtained fresh from three En(a−) individuals are indeed relatively resistant to invasion by *P. falciparum* although they are able to support parasite development. Our results also indicate that these cells may be a useful model in the search for a putative receptor for *P. falciparum*.

We have compared the rates of invasion by *P. falciparum* of normal and En(a−) cells obtained from the British[9] (MEP), Finnish[10] (GW) and Canadian[11] (RL) homozygous En(a−) individuals using an *in vitro* culture system[6–8] (Table 1). The numbers of parasites found in the En(a−) cells were significantly less than in the En(a−) controls using either the continuous asynchronous culture system[6] (incubation period 72 h) or the synchronized short-term (12–18 h) microtissue culture tray system[8]. The fact that parasite numbers are smaller in the longer-term cultures with cells from MEP is probably explained by some parasites being able to go through more than one schizogony, thus amplifying the differences between the En(a−) and En(a+) cells.

Because the reduction of invasion into En(a−) cells was relative rather than absolute, it was particularly important to control for contamination by En(a+) cells unavoidably present in the infected blood. This was achieved by using fetal haemoglobin (HbF) as a marker of En(a+) cells in the parasite culture. By using a modification of the acid elusion technique[12] to stain for HbF, to exclude En(a+) cells from parasite counts[13], we confirmed that there was some invasion of En(a−) cells but much less than of En(a+) cells. The more marked reduction in parasite numbers in GW and RL compared with MEP is surprising because whereas MEP cells, in addition to being deficient in glycophorin A, are also partially deficient in glycophorin B[14], a glycoprotein sharing considerable homology with glycophorin A[15], the cells of GW and RL lack only glycophorin A.

Once inside the En(a−) cells, parasites were able to develop as successfully as in En(a+) cells (Table 2). In these experiments also, HbF was used as a marker to control for contaminating En(a+) cells from the infected blood. These results indicate that defective invasion rather than development is the mechanism whereby parasites fail to interact with En(a−) cells.

Because of previous sensitization by transfusion or during pregnancy, the sera of En(a−) individuals usually contain anti-En[a] antibodies of varying titres. Just as En(a−) cells have several membrane alterations, anti-En[a] sera contain antibodies with at least three specificities—against the portions of the glycophorin A molecule which are respectively sensitive and insensitive to trypsin and against the Wr[b] antigen[5]. The effect of anti-En[a]

Table 1 Invasion of En(a+) and En(a−) cells by *P. falciparum*

Test cells	Incubation time (h)	Parasites per 100 red cells En(a+)	En(a−)	En(a−)/En(a+)
MEP_1	72	8.4	1.4	0.17
		6.2	2.0	0.32
		4.9	1.9	0.39
				Mean 0.29 ± 0.06
MEP_2	12–18	7.5	3.2	0.43
		12.0	5.0	0.43
		10.3	5.9	0.57
				Mean 0.48 ± 0.05
GW	12–18	13.4	1.0	0.07
		7.2	0.6	0.08
				Mean 0.08 ± 0.07
RL	12–18	6.7	0.9 (1.3)	0.13
		9.3	1.4 (2.1)	0.15
		7.5	1.1 (2.2)	0.15
				Mean 0.14 ± 0.01

The Uganda–Palo Alto strain of *P. falciparum* used was maintained in continuous culture using previously described methods[6–8]. In the 72-h cultures, cells containing asynchronous parasites were mixed in a 1 : 10 dilution with uninfected En(a+) or En(a−) cells and grown in Petri dishes with changes of medium every 12 h. In all cases, the cells, if not used immediately, were stored at 4 °C in citrate/phosphate/dextrose (CPD). After 72 h, samples from the cultures were smeared and stained to count the number of parasites (all stages) present. At least 1,000 cells were counted. The starting parasite count in these experiments was ~1%. Parasite growth for the short-term invasion assay was synchronized using a 5% sorbitol solution[20]. Schizonts were concentrated with a 3% gelatin solution (Plasmagel)[21] and mixed with uninfected cells at a dilution of 1 : 6–1 : 20 to obtain a 1.5% cell suspension with 2–5% of the cells containing schizonts; 200 µl of the mixture were pipetted into each well of a microtissue culture plate, incubated for 12–18 h at 37 °C in a mixture of 5% O_2/7% CO_2/88% N_2 and then smears were made and stained. Parasite numbers were assessed by counting the number of ring forms in at least 1,000 cells. The relative multiplication in the En(a−) cells was calculated as a proportion of the controls ± s.e.m. Cells from En(a−) individuals were obtained on two occasions from the Sheffield Blood Transfusion Service, UK (MEP), from the Red Cross Blood Transfusion Service, Helsinki (GW) and from the Red Cross Blood Transfusion Service, Montreal (RL). These and control En(a+) bloods taken at the same time were collected in CPD or acid/citrate/dextrose (ACD). (In the experiments with the En(a−) cells of RL, infected blood came from cultures containing fetal haemoglobin (HbF). Thus by staining smears using a modification of the acid elution technique[12] the counting of parasites in the HbF-containing En(a+) cells could be excluded. Figures in parentheses show parasite counts in all cells (adult and HbF-containing).

Table 2 Development of *P. falciparum* in En(a−) cells

	Invasion Parasites per 100 red cells (En(a−)/En(a+))	Development Schizonts per 100 singly infected cells En(a+)	En(a−)	(En(a−)/En(a+))
	0.13*	79	78	0.99
	0.15*	80	79	0.99
	0.15*	80	82	1.03
	0.35	68	70	1.03
	0.32	67	70	1.04
	0.16	70	77	1.10
	0.42	71	79	1.11
Mean	0.24 ± 0.04			1.04 ± 0.02

The methods used were as for Table 1. The cells used were from the Canadian En(a−) individual (RL). The medium in the microtissue culture wells was changed at ~15 and 40 h. Development was assessed at 48 h by counting the number of parasites in 100 singly infected cells that had matured to the schizont stage (that is, a large parasite with the chromatin clearly divided into at least two distinct entities).

* In these experiments infected blood came from cultures containing HbF.

0028-0836/82/180064—03$01.00

Table 3 Invasion of En(a+) cells by *P. falciparum* in the presence of anti-En^a serum

Serum	Titre (reciprocal)	ABS	Parasites per 100 red cells Anti-En^a (unabsorbed)	Anti-En^a (absorbed)	Unabsorbed / Absorbed
GW_1	64	9.2	1.7	Not tested	
		5.6	0.6	5.9	0.10
		8.2	1.1	9.3	0.12
		7.4	1.0	7.9	0.13
		3.8	0.6	3.9	0.15
					Mean 0.13 ± 0.07
GW_2	8	10.1*	4.0	12.7	0.31
		11.5	6.2	14.4	0.43
		9.2*	4.7	10.6	0.44
		11.6	5.0	9.1	0.55
		17.4	12.2	16.0	0.76
					Mean 0.50 ± 0.08
MEP	8	16.1	6.4	18.9	0.34
		7.7	3.5	9.4	0.37
		9.2*	4.4	9.5	0.46
		11.2	5.6	11.0	0.51
		7.6	4.6	8.2	0.56
		10.1*	5.4	9.5	0.59
		17.4	8.7	14.4	0.60
					Mean 0.49 ± 0.04

The methods used were as for Table 1 except that 50 µl of the medium per well was substituted with 50 µl of the serum to be tested[22]. Group O-positive cells were used in all experiments. Sera from the En(a−) individuals were obtained on several occasions. The first example from GW(GW_1) contained sodium azide added as a preservative. As azide is toxic to parasites, this serum was dialysed with parallel AB serum (ABS) controls against three changes of 1 l of phosphate-buffered saline *p*H 7.4. The sera from MEP and GW_2 were taken directly from the patients then placed in sterile preservative-free containers. The titres of the sera were determined by indirect anti-immunoglobulin tests carried out on all the sera at the same time. The sera possessing anti-En^a antibodies were absorbed by incubating them with an equal volume of group O +ve cells for 1 h at 37 °C. Similar treatment of ABS did not alter the rates of invasion (data not shown). Successful absorption of these anti-En^a sera was confirmed by indirect antiglobulin testing using En(a+) cells.

* In these experiments the concentration of serum was increased from 50 to 100 µl per well.

sera on parasite multiplication was examined in several experiments (Table 3). The presence of antibodies on the red cell surface was confirmed by direct anti-immunoglobulin testing. There was a significant reduction in the number of parasites in cultures to which sera were added, regardless of the method of collection, either compared with ABS or, more importantly, with the same anti-En^a sera after absorption with En(a$^+$) cells. The greater inhibition of invasion using the serum of GW correlated with its titre in an indirect antiglobulin test. The other sera used were of relatively low titre, having been taken several years after the individuals were sensitized.

To exclude the possibility that coating of red cells with anti-En^a antibody might reduce parasite invasion in a non-specific manner, various other sera directed against commonly occurring antigens present on red cells were examined in the same way. Two of these antibodies were directed against Rhesus antigens (anti-C and anti-c), one against the Kidd system (anti-Jk^a) and another against a Duffy determinant (anti-Fy^a). In no case was the number of parasites found in cells coated with antibody significantly less than in the controls (Table 4). On the other hand, there was a significant decrease in parasite numbers when sera from two patients who had recently recovered from cerebral malaria were added to the cultures. Anti-immunoglobulin tests confirmed the presence of antibody on the surface of the cells incubated with antisera to red cell antigens, but not on the cells incubated with the patients' sera. Antibodies in the latter are thought to act via merozoite agglutination rather than by receptor blockade[16].

Recent structural analysis of glycophorin A indicates that the M and N antigens are determined by its five amino-terminal amino acids and three oligosaccharides[17]. We further examined whether this most external portion of the glycophorin A molecule might be involved in the interaction with the parasite by comparing the relative rates of invasion of *P. falciparum* into cells bearing the M, the N and both antigens. The rates of invasion observed in MN, MM or NN cells were similar, indicating that the merozoite was unable to distinguish between these two antigens (Table 5).

These experiments have shown a significant reduction in invasion of En(a−) cells by *P. falciparum*. This may indicate either that there is a specific receptor site on the major sialoglycoprotein, or that there is a non-specific effect due to rearrangement of the membrane proteins in the absence of this major component. However, the presence of anti-En^a antibody caused

Table 4 Invasion of En(a+) cells by *P. falciparum* in the presence of antisera other than anti-En^a

Strain of parasite	ABS (control)	Parasites per 100 red cells Anti-C	Anti-c	Anti-JK^a	Anti-Fy^a	JG	DH
A	8.2	7.7	8.4	8.2	7.7	4.2	0.0
A	3.8	3.6	4.1	4.0	3.7	0.0	0.0
B	5.7	6.5	7.0	6.5	6.6	NT	NT
C	9.1	7.9	7.8	5.5	8.8	4.5	0.3
C	5.6	2.0	5.3	4.1	5.3	4.8	2.3
	Mean $\frac{\text{Test}}{\text{Control}}$	0.85 ± 0.13	1.05 ± 0.05	0.96 ± 0.08	0.99 ± 0.06	0.47 ± 0.18	0.11 ± 0.10

The methods used were as for Tables 1 and 3. Three strains of parasite were used: A, Uganda–Palo Alto; B, wild West African; and C, wild East African. The wild strains were obtained from infected patients returning to Oxford. Sterile, preservative-free antisera were obtained from the Oxfordshire Regional Blood Transfusion Service. Sera were also obtained, at least 1 month after completing chloroquine therapy, from two patients convalescing after severe *P. falciparum* infections (JG and DH). The relevant typing of the En(a+) cells used was as follows: ABO group O; Rh-positive (C$^+$, D$^+$, E$^+$, c$^+$, e$^+$); M+, N+, Fy(a+), Fy(b+); Jk(a+), Jk(b−), Wr(a−), Wr(b+).

NT, not tested.

Table 5 Invasion by *P. falciparum* of cells expressing either the M or N or both antigens

	Parasites per 100 red cells	
MN cells (Control)	MM cells (Test)	NN cells (Test)
8.8	9.1	9.8
7.5	6.6	7.8
6.4	6.7	6.5
Mean $\left(\frac{\text{Test}}{\text{Control}}\right)$	0.99 ± 0.05	1.06 ± 0.03

The methods were as for Table 1. The cells were obtained in CPD from the Oxfordshire Regional Blood Transfusion Service.

a significant reduction in parasite numbers in En(a+) cells which was prevented by previous absorption with En(a+) cells. This effect was not seen when other antibodies were used to coat the red cells. This finding suggests that the relationship between the parasite and the missing membrane component of the En(a−) cells is more specific. The fact that anti-En^a antibody (GW_2 and MEP) failed to reduce invasion of normal cells to the level found in En(a−) cells may have been due to low antibody titres resulting in insufficient blockade of this high density receptor (glycophorin A), of which there are at least 500,000 copies per cell[18]. Moreover the reduction in invasion caused by the low titre anti-En^a sera cannot exclude the role of the Wr^b antigen which is also recognized by this serum. Although these antisera could act by damaging the mature schizont-infected cell, by merozoite agglutination, or by premature lysis of infected En(a−) cells, these data, together with the reduced invasion of En(a−) cells, are more consistent with the anti-En^a sera acting via receptor blockade.

Of the many types of red cells that have been examined previously, only those containing haemoglobin S or predominantly haemoglobin C have shown rates of invasion inappropriate to their metabolic age[19]. En(a−) cells must now be added to this list. Miller has shown that trypsin-treated red cells are relatively resistant to invasion by *P. falciparum*[4]. Glycophorin A is one of the major targets for cleavage by this enzyme. These findings, together with the present studies, suggest that glycophorin could serve as a receptor for *P. falciparum*. The development of monoclonal antibodies against specific regions of these glycoproteins should allow this possibility to be explored further.

We thank Dr W. Wagstaff, Anna Pirkola, Dr R-M. Guévin, Mr V. Taliano and Mr M. Woodward for providing the numerous samples of cells, sera and reagents. We also thank Drs R. Sanger and P. Tippett for assistance. G.P. was supported by the Wellcome Trust and Rockefeller Foundation.

Note added in proof: Since submission of this letter, Perkins[23] has shown that glycophorin A in solution and Fab′ fragments of antibody against glycophorin A will inhibit merozoite invasion.

Received 22 April; accepted 1 August 1981.

1. McGhee, R. B. *Ann. N.Y. Acad. Sci.* **56,** 1070–1073 (1953).
2. Butcher, G. A., Mitchell, G. H. & Cohen, S. *Nature* **244,** 40–42 (1973).
3. Miller, L. H., Dvorak, J., Shiroishi, T. & Durocher, J. *J. exp. Med.* **138,** 1597–1601 (1973).
4. Miller, L. H. *et al. J. exp. Med.* **146,** 277–281 (1977).
5. Anstee, D. J. *Sem. Hemat.* **18,** 13–31 (1981).
6. Trager, W. & Jensen, J. B. *Science* **193,** 673–675 (1976).
7. Haynes, J. D., Diggs, C. L., Hones, F. A. & Desjardins, R. E. *Nature* **263,** 767–769 (1976).
8. Phillips, R. S., Trigg, P. I., Scott-Finnigan, T. J. & Bartholomew, R. K. *Parasitology* **65,** 525–535 (1972).
9. Darnborough, J., Dunsford, I. & Wallace, J. A. *Vox. Sang.* **17,** 241–255 (1969).
10. Furuhjelm, U. *et al. Vox. Sang.* **17,** 256–278 (1969).
11. Taliano, V. *et al. Vox. Sang.* **38,** 87–93 (1980).
12. Kleihauer, E., Braun, H. & Betke, K. *Klin. Wochenschr.* **35,** 637–638 (1957).
13. Diggs, C. L. *et al. J. Parasit.* **57,** 187–188 (1971).
14. Metaxas, M. N. & Metaxas-Buehler, M. in *Human Blood Groups*, 344–352 (Karger, Basel, 1976).
15. Furthmayr, H. *Nature* **271,** 519–524 (1978).
16. Miller, L. H., Aikawa, M. & Dvorak, J. A. *J. Immun.* **114,** 1237–1242 (1975).
17. Wasniowska, K., Drzeniek, Z. & Lisowska, E. *Biochem. biophys. Res. Commun.* **76,** 385–390 (1977).
18. Steck, T. L. *J. Cell Biol.* **62,** 1–19 (1974).
19. Pasvol, G., Weatherall, D. J. & Wilson, R. J. M. *Br. J. Haemat.* **45,** 285–295 (1980).
20. Lambros, C. & Vanderberg, J. P. *J. Parasit.* **65,** 418–420 (1979).
21. Pasvol, G., Wilson, R. J. M., Smalley, M. E. & Brown, J. *Ann. trop. med. Parasit.* **72,** 87–88 (1978).
22. Wilson, R. J. M. & Phillips, R. S. *Nature* **263,** 132–134 (1976).
23. Perkins, M. *J. Cell Biol.* **90,** 563–567 (1981).

Virus-induced diabetes mellitus: autoimmunity and polyendocrine disease prevented by immunosuppression

Takashi Onodera, Usha R. Ray, Kathleen A. Melez, Hoshibumi Suzuki, Antonio Toniolo & Abner Louis Notkins

Laboratory of Oral Medicine, National Institute of Dental Research, National Institutes of Health, Bethesda, Maryland 20205, USA

Autoantibodies to cytoplasmic and surface antigens on pancreatic islet cells have been found in many newly diagnosed cases of insulin-dependent diabetes mellitus (IDDM)[1-9]. Various autoantibodies (for example, against islet cells, adrenals, thyroid, gastric mucosa and intrinsic factor) have also been found in patients with polyendocrine disease[1,2,8,9]. What triggers the production of these antibodies and their importance in the pathogenesis of these diseases are still unclear. Recently, we found in an animal model for studing polyendocrine disease[10] that reovirus type 1 (reo-1) infected cells in both the pancreas and pituitary, and produced a disease characterized by hyperglycaemia and retarded growth[10]. Moreover, autoantibodies that reacted with antigens in pancreatic islets, the anterior pituitary and gastric mucosa were found, one of which reacted with insulin and another with growth hormone (GH). Using recombinant viruses[11] we showed that the *S1* gene segment from reo-1 was required for the induction of autoantibodies to GH. We report here our search for additional autoantibodies, especially against surface antigens, and our efforts to evaluate the pathological contributions of the lytic effect of the virus on cells compared with the immunopathological response of the host. Our experiments showed that reo-1 induced autoantibodies that reacted with antigens on the surface of islet cells, thymocytes and a GH-producing cell line. In addition, we found that treatment with immunosuppressive agents depressed autoantibody production and prevented development of the polyendocrine disease.

Except where indicated, male and female SJL mice were infected intraperitoneally (i.p.) at 5–7 days of age with the Lang strain of reo-1 that had been passaged 10–14 times in pancreatic β-cell cultures[10]. Rabbit anti-mouse lymphocyte serum (ALS) and rabbit anti-mouse thymocyte serum (ATS) (MA Bioproducts, Walkersville, Maryland), with cytotoxic titres 1:6,400 and 1:1,600 respectively, and cyclophosphamide (Cy) (20 mg per kg)[12] were used as immunosuppressive agents. Single-point glucose tolerance tests (GTTs) were performed 60 min after i.p. injection of 2 mg glucose per body weight[10]. Plasma GH was measured by radioimmunoassay[10,13], and antibody to reo-1 was determined by a standard haemagglutination-inhibition technique[14]. Indirect immunofluorescence was carried out as described previously[10].

Three different immunosuppressive agents were tested for their effect on the development of reovirus-induced diabetes. As shown in Fig. 1, reo-1 produced abnormal GTTs in SJL mice, but these abnormalities were reduced or prevented by treatment with ALS (Fig. 1*a*). Similarly, ATS and Cy prevented the development of reovirus-induced hyperglycaemia (Fig. 1*b*). Reo-1 also produced some glucose abnormalities in NFS mice and these abnormalities were suppressed by ALS (Fig. 1*c*). All eight experiments performed with reo-1 and immunosuppressive agents revealed similar trends, although in some experiments the reovirus-induced abnormal GTTs were much less marked than those observed in Fig. 1*a* and *b*.

Immunosuppression prevented not only the development of hyperglycaemia but also growth retardation. Table 1 shows that

virus-infected immunosuppressed mice gained weight at almost the same rate as uninfected controls, and that oily hair and steatorrhoea were greatly decreased by ALS. Mortality was also markedly reduced. Moreover, the infected mice with retarded growth had considerably lower plasma GH levels than the infected and immunosuppressed mice.

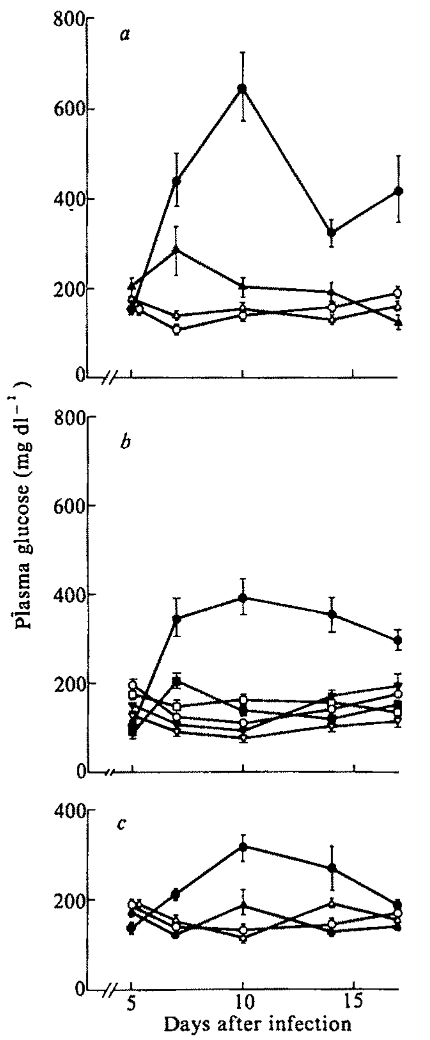


Fig. 1 Effect of immunosuppressive agents on glucose tolerance tests in reo-1-infected mice. Mice were infected with 5×10^5 plaque-forming units (PFU) of virus and single-point glucose tolerance tests were performed at different times thereafter. Immunosuppressive treatment consisted of three 0.1-ml i.p. injections of ALS or ATS on days 0, 4 and 7 after infection or a single i.p. dose of Cy 1 day after infection. 10 to 35 mice were tested at each time point in each group. The mean ± s.e.m. is represented by symbols and vertical lines. *a*, SJL mice: ●, reovirus-infected; ○, uninfected; ▲, infected and treated with ALS; △, uninfected and treated with ALS. *b*, SJL mice: ●, reovirus-infected; ○, uninfected; ■, infected and treated with ATS; □, uninfected and treated with ATS; ▼, infected and treated with Cy; ▽, uninfected and treated with Cy. *c*, NFS mice: ●, reovirus-infected; ○, uninfected; ▲, infected and treated with ALS; △, uninfected and treated with ALS.

To see whether the immunosuppressive agents might be preventing clinical symptoms by inhibiting the autoimmune response, we measured the titre of autoantibodies in the sera of infected mice reacting with tissue sections from uninfected normal mice[10]. As shown in Table 2, autoantibodies to the anterior pituitary, islet cells and GH_3 cells (a rat cell line which produces both GH and prolactin)[15] were greatly reduced, or not detected, in sera from reovirus-infected immunosuppressed mice. Immunosuppression also reduced the percentage of mice with autoantibodies to GH and insulin.

In addition to autoantibodies reacting with cytoplasmic antigens, autoantibodies to surface antigens on viable islet cells, GH_3 cells and thymocytes, but not mouse fibroblasts, were found by avidin-biotin immunofluorescence[16]. In our hands, this method is more sensitive than standard indirect immunofluorescence. Nonetheless, the titre of these antibodies was quite low and they reacted with only a small number (10–15%) of islet cells, GH_3 cells and thymocytes. In reovirus-infected immunosuppressed mice, these autoantibodies could not be detected. Immunosuppression also reduced, but did not eliminate, the antibody response to reo-1 (Table 2); this may account for the ability of the immunosuppressed animal to cope with the viral infection.

As previously reported, reovirus infection produced mild β-cell damage and insulitis, with infiltration of mononuclear and plasma cells[10]. In the present experiments, immunosuppression with ALS greatly reduced the degree of insulitis and β-cell damage (data not shown). Precisely how the viral infection triggers the autoimmune and inflammatory responses[10] and what role, if any, autoantibodies to thymocytes have in this process are unclear.

The possibility that the immunosuppressive agents were preventing the polyendocrine diseases by inhibiting virus replication (such as by antiviral antibody in the immunosuppressive sera or induction of interferon), rather than by suppressing the autoimmune response, was evaluated by measuring virus growth *in vivo*. Figure 2 shows that the former possibility is unlikely as virus titres in homogenates of liver, heart and pancreas were approximately the same in untreated and ALS-treated mice at various times after infection.

Viral infections cause tissue damage by two principal mechanisms: they can directly lyse cells as a result of virus replication, and they can trigger in the host an immunopathological response[17–19]. In the case of reovirus, both mechanisms may operate; infected cells are certainly destroyed as a result of virus replication and a recent study suggests that some cells may also be destroyed by cytolytic T lymphocytes specific for reovirus antigens on the surface of cells[20].

The present experiments suggest another possible mechanism by which reovirus might cause disease: the infection triggers the production of autoantibodies that react with antigens on the surface of normal uninfected cells or with cell products (for example hormones). These interactions could act pathologically by directly destroying hormone- producing cells or by binding to and reducing the concentration of hormones in the circulation.

In experimental animals, we previously showed that neither

Table 1 Reversal of reovirus-induced growth retardation, clinical symptoms and mortality by immunosuppression

Groups		Weight (g)		% Oily hair and steatorrhoea		% Mortality		GH (ng ml^{-1})
Infection	ALS	SJL	NFS	SJL	NFS	SJL	NFS	SJL
+	–	7.6±2.3*	8.3±2.5*	63	50	43	40	16±8
+	+	12.3±3.5†	13.7±1.8†	0	0	25	13	48±4
–	+	11.4±1.0†	13.6±1.0†	0	0	0	0	ND
–	–	14.1±2.4	14.9±2.2	0	0	0	0	41±6

SJL and NFS mice were infected and treated with ALS as in Fig. 1. Each group consisted of 10–20 mice. Body weights given were for 17 days after infection (mean±s.d.), the incidence of oily hair and steatorrhoea was determined 10 days after infection, cumulative mortality was assessed at 17 days after infection, and plasma GH 14 days after infection (mean±s.e.m.). ND, not done.

* $P < 0.001$ compared with uninfected and untreated controls.

† Not statistically significant compared with uninfected and untreated controls.

Table 2 Effect of immunosuppression on autoantibody titres in reovirus-infected mice

Groups		Autoantibodies to: Cytoplasmic antigens*			Surface antigens†			Hormones‡		
Infection	ALS	Anterior pituitary	Islets	GH_3	Islets	GH_3	Thymocytes	Insulin	GH	Antibody to reovirus
+	−	32	8	16	4	2	8	56	89	256
+	+	2	2	<2	<2	<2	<2	16	20	16
−	+	<2	<2	<2	<2	<2	<2	0	0	<8
−	−	<2	<2	<2	<2	<2	<2	0	0	<8

Reo-1 infection and ALS treatment were as in Fig. 1. Autoantibodies were measured in pooled sera from 10 mice. Similar results were obtained with pooled sera from two other experiments. Titres for reovirus antibody are given as reciprocal antibody titres by haemagglutination-inhibition tests to reo-1 in pooled sera 17 days after infection.

* Bouin's-fixed sections of anterior pituitary and pancreas from uninfected mice, or acetone-fixed rat GH_3 cells were incubated with serial dilutions of mouse sera (obtained 14 days after infection) and stained with fluorescein-labelled anti-mouse IgG. Titres represent the reciprocal of the highest dilution of sera giving positive fluorescence[10].

† Suspensions of cultured mouse SJL pancreatic islet cells, or rat GH_3 cells, or freshly prepared thymocytes from weanling SJL mice were incubated with pooled mouse sera obtained 14 days after infection and stained with biotin-conjugated anti-mouse immunoglobulin followed by fluorescein-labelled avidin[16]. Titres represent the reciprocal of the highest dilution of sera giving positive fluorescence.

‡ Individual sera from 10–20 mice per group (obtained 17 days after infection) were tested for antibodies to insulin and GH by an enzyme-linked immunosorbent assay[10]. Sera with titres >10 were considered positive and numbers represent per cent positive.

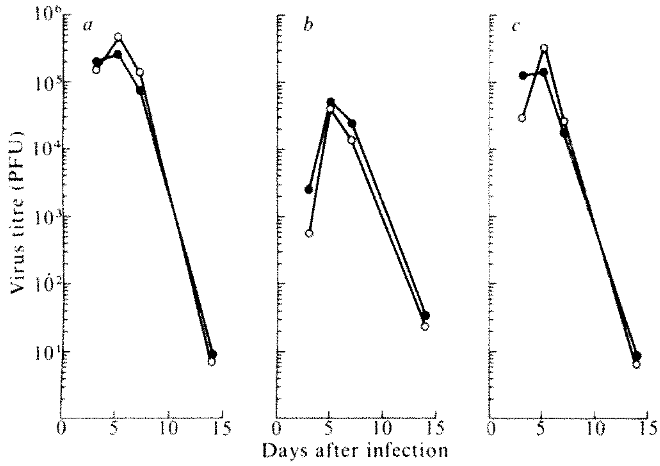


Fig. 2 Reovirus titre in ALS treated mice. SJL mice were infected with 5×10^5 PFU and treated with ALS as in Fig. 1. Virus titres in homogenates of liver (*a*), heart (*b*) and pancreas (*c*) were measured at different times after infection. Titrations were performed on L929 cells by a standard plaque assay[10] and expressed as PFU per organ. Each point represents the mean titre of two to four individual mice. ○, Infected mice given ALS; ●, infected mice not ALS.

virus alone nor sub-diabetogenic doses of streptozotocin alone could produce hyperglycaemia, whereas treatment with both agents resulted in overt diabetes[21]. Our interpretation was that neither virus nor streptozotocin alone depleted the β-cell reserve sufficiently to result in diabetes. Reovirus-induced polyendocrine disease may also be the result of cumulative injury. Both the direct lytic effect of the virus and autoimmune pathology may be required to decrease the hormone reserve sufficiently to make the disease overt. The possibility that virus-induced autoimmunity contributes to the pathogenesis of other diseases merits investigation.

We thank Tecolia Brown and Royd Ellis for technical help and Edith Rian for editorial assistance.

Received 7 January; accepted 15 March 1982.

1. Bottazzo G. F., Florin-Christensen, A. & Doniach, D. *Lancet* **ii**, 1279–1282 (1974).
2. MacCuish, A. C., Irvine, W. J., Barnes, E. W. & Duncan, L. J. P. *Lancet* **ii**, 1529–1531 (1974).
3. Irvine, W. J., Gray, R. S. & McCallum, C. J. *Lancet* **ii**, 1097–1102 (1976).
4. Lendrum, R., Walker, G. & Gamble, D. R. *Lancet* **i**, 880–883 (1975).
5. Doniach, D. & Bottazzo G. F. in *Pathobiology Annual* (ed. Joachim, H. L.) 327–346 (Appleton-Century-Crofts, New York, 1977).
6. Lernmark, A. *et al. New Engl. J. Med.* **299**, 375–380 (1978).
7. Dobersen, M. J., Scharff, J. E., Ginsberg-Fellner, F. & Notkins, A. L. *New Engl. J. Med.* **303**, 1493–1498 (1980).
8. Cahill, G. F. & McDevitt, H. O. *New Engl. J. Med.* **304**, 1454–1465 (1981).
9. Doniach, D. & Bottazzo, G. F. in *Clinical Immunology Update* (ed. Franklin, E. D.) 96–121 (Elsevier, Amsterdam, 1981).
10. Onodera, T. *et al. J. exp. Med.* **153**, 1457–1473 (1981).
11. Weiner, H. L., Drayana, D., Averill, D. R. Jr & Fields, B. N. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5744–5748 (1977).
12. Openshaw, H., Asher, L. V. S., Wohlenberg, C., Sekizawa, T. & Notkins, A. L. *J. gen. Virol.* **44**, 205–215 (1979).
13. Birge, C. A., Peake, G. T., Mariz I. K. & Daughaday, W. H. *Endocrinology* **81**, 195–204 (1967).
14. Weiner, H. L., Powers, M. L. & Fields, B. N. *J. infect. Dis.* **141**, 609–616 (1980).
15. Bancroft, F. C. & Tashjian, A. H. Jr *Expl Cell Res.* **64**, 125–128 (1971).
16. Ledbetter, J. A., Rouse, R. V., Micklem, H. S. & Herzenberg, L. A. *J. exp. Med.* **152**, 280–295 (1980).
17. Notkins, A. L. (ed.) *Viral Immunology and Immunopathology* (Academic, New York, 1975).
18. Oldstone, M. B. A. *Prog. med. Virol.* **19**, 84–119 (1975).
18. Zinkernagel, R. M. *Compreh. Virol.* **15**, 171–204 (1979).
20. Finberg, R., Weiner, H. L., Fields, B. N., Benacerraf, B. & Burakoff, S. J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 442–446 (1979).
21. Toniolo, A., Onodera, T., Yoon, J. W. & Notkins, A. L. *Nature* **288**, 383–385 (1980).

Attenuated reovirus type 3 strains generated by selection of haemagglutinin antigenic variants

Dale R. Spriggs* & Bernard N. Fields*†

* Department of Microbology and Molecular Genetics, Harvard Medical School, and † Department of Medicine, Division of Infectious Diseases, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA

Infection of suckling mice with mammalian reovirus type 3 results in fatal encephalitis associated with marked destruction of neurones[1–4]. The viral tropism specific to neural cells is determined by the reovirus haemagglutinin[3,4]. This haemagglutinin (HA), the $\sigma 1$ polypeptide, is encoded in the S1 dsRNA segment[5,6]. In addition to determining cell and tissue tropism, the viral haemagglutinin determines humoral and cellular immune specificity[7–10], binding to cellular microtubules[11] and inhibition of cellular DNA synthesis[12]. To elucidate the structural basis for the functions of the reovirus type 3 HA, we have recently isolated and characterized several anti-HA monoclonal antibodies, which we have used to show that there are distinct functional domains on the reovirus type 3 HA[13]. One group of monoclonal antibodies neutralizes viruses while a second inhibits haemagglutination. Thus there are distinct HA functional domains involving haemagglutination and neutralization. One of the neutralizing monoclonal antibodies was used to select reovirus antigenic variants that were no longer neutralized by the monoclonal antibodies. These 'variant' viruses were tested for their neurovirulence by determining their capacity to replicate in mouse brains, leading to fatal encephalitis. We report here that all variants are markedly less virulent then the parental type 3 virus from which they were selected. These findings demonstrate directly the critical role of the viral haemagglutinin in neurovirulence and show that variant viruses that are altered at a site recognized by neutralizing antibody cannot grow efficiently in neural tissue.

0028–0836/82/180068—03$01.00

Reovirus serotype 3 viruses with antigenically altered HA proteins were selected by incubating the neurovirulent virus stock (Dearing strain) with excess neutralizing monoclonal antibody (designated G5). Viral plaques that appeared on the plates were isolated and passaged a second time in the presence of G5 antibody to eliminate clones that were not resistant to the antibody. Resistant viruses (variants) were then grown in mouse L cells to give virus stocks. Three variant viruses (A, F and K) isolated in this manner were tested for their relative ability to resist neutralization by the G5 monoclonal antibody (Table 1). As reported previously, the parental reovirus type 3 (Dearing) strain was neutralized efficiently by G5 (ref. 13). All three variants showed an altered response to G5 antibody: the F and K variants were totally resistant to the antibody while the A variant was neutralized to a level intermediate between that of the Dearing parent and the F and K variants. For example, the A variant was neutralized at 1:20 and 1:100 dilutions of G5, but the plaque reduction at 1:500 was of borderline significance. The Dearing strain was neutralized at all the dilutions tested. Previous studies have shown that the G5 antibody neutralizes the Dearing strain at dilutions up to 1:12,500 (ref. 13). In contrast to the results with the G5 monoclonal antibody, the three variants were indistinguishable from the parental reovirus type 3 Dearing virus in neutralization tests using a type-specific hyperimmune antiserum (Table 1). These data indicate that all three variants encode an haemagglutinin having antigenic alterations that are readily detected with monoclonal antibodies but not with a standard hyperimmune antiserum.

To determine if the variants were altered in terms of virulence, we compared them with regard to their relative ability to cause fatal neurological disease in suckling mice. The type 3 Dearing parent is highly neurovirulent; as few as 10 plaque-forming units (PFU) inoculated intracerebrally (i.c.) in suckling mice kill 50% of mice by 14 days while higher doses kill 100% (ref. 3; Table 2). To compare the neurovirulence of type 3 Dearing with the variant viruses, we injected varying doses of virus into suckling mice and determined the survival pattern for each infected group of animals (Table 2). The F and K variants had LD_{50}s (lethal dose for 50% of the population) values of 3×10^7 and $>3\times10^7$ PFU, respectively, while the A variant had an intermediate LD_{50} of 1.8×10^5 PFU. These results indicate that all the variants are at least 10,000 times less neurovirulent than the Dearing strain from which they were derived. Interestingly, the A variant that was intermediate in antigenic reactivity, also exhibited an intermediate LD_{50} (1.8×10^5) whereas the F and K variants, which were not neutralized by G5, showed the greatest reduction in virulence. These data suggest that the loss of antigenic reactivity with the G5 antibody is directly related to the neurovirulence.

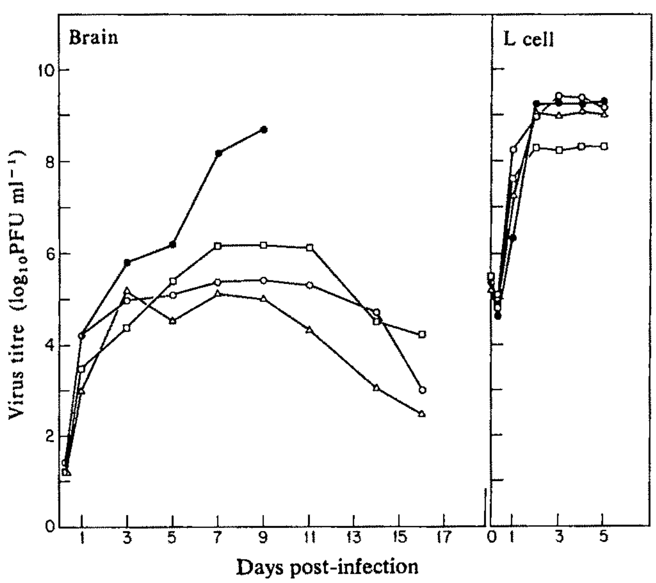


Fig. 1 Growth patterns of reovirus type 3 (Dearing, ●) and antigenic variants A (□), F(○) and K (△) in mouse brains (left) and mouse L cells in tissue culture (right). Virus growth in mouse brains was determined by injecting all mice i.c. with 1×10^3 PFU of virus. Three mice were killed at each time point and the virus titres in the individual brains were determined by titration of brain homogenates on mouse L cells at 37 °C as described previously[4]. Each value represents the average of three titrations. Virus growth in mouse L cells was determined by infecting monolayers of 2×10^5 cells at a multiplicity of infection of 5 PFU per cell. Triplicate samples were taken for titration at each indicated time point[16].

Table 1 Neutralization of reovirus serotype 3 (Dearing) and HA antigenic variants by monoclonal antibody and hyperimmune antiserum

Antibody	Dilution of antibody	% Inoculum resistant to antibody: Dearing strain	Variant strains: A	F	K
G5	1:20	0	21	100	88
	1:100	0	18	100	100
	1:500	5	48	100	100
Hyperimmune	1:200	6	7	4	3
	1:400	5	4	2	4
	1:800	14	4	5	9

Neutralization tests were performed as described previously[13]. Briefly, dilutions of antibody were incubated with 100 PFU of virus at 34 °C for 1 h. These mixtures were then inoculated on to mouse L cells. The G5 antibody was used to select the variants and was used in these experiments as a purified IgG fraction[13]. Values represent the percentage of plaques observed in the presence of antibody relative to a phosphate-buffered saline control. The hyperimmune antiserum is a reference goose anti-reovirus type 3 antiserum obtained from the National Institute of Allergy and Infectious Diseases (V703-501-570).

Because neurovirulence is correlated with the ability of reovirus type 3 to replicate in suckling mouse brains, we analysed the growth of type 3 reovirus and the three variants in mouse brains (Fig. 1). The reovirus type 3 (Dearing) virus grows to high titres in the brain before death of the infected mice[3,4]. The virus yield in these experiments was $\sim1\times10^9$ PFU ml^{-1}, but titres were often $\geq1\times10^{10}$ PFU ml^{-1}. The three variants differed significantly from the Dearing strain in their ability to grow in mouse brains during the acute infection, particularly after the fifth day post-inoculation (Fig. 1). The F and K variants showed most reduction in growth while the A variant was again intermediate between the F and K variants and the Dearing virus.

To determine whether the inability of the three variants to grow in brains was due to a general defect of growth or was

Table 2 Neurovirulence of reovirus serotype 3 (Dearing) and HA antigenic variants A, F and K

Inoculum (PFU)	% Mice alive on day 14 post-infection: Dearing strain	Variant strains: A	F	K
1×10^1	50	100	100	100
1×10^2	0	100	100	100
1×10^3	0	100	100	100
1×10^4	NT	100	100	100
1×10^5	NT	90	100	100
3×10^5	NT	0	100	100
3×10^6	NT	0	90	100
3×10^7	NT	0	50	90
LD_{50}	10^1	1.8×10^5	3×10^7	$>3\times10^7$

Groups of CD mice (24–48 h old) were injected i.c. with varying doses of virus and observed for 14 days[3,4]. Animals began to die between days 8 and 10 in most cases. Values represent the percentage of mice alive at day 14 post-infection. NT, not tested. The LD_{50} is the dose of virus required to kill 50% of the animals in each group by day 14. LD_{50} values were calculated by the Reed and Muench method[15].

specific to growth in the brain, we examined the replication cycle of the variants and Dearing virus in mouse L cells (Fig. 1). Both reovirus type 3 Dearing and the three variants grew to a high titre. The variants were also not temperature-sensitive for growth in L cells (data not shown). Examination of brains infected with variants A, F and K showed a marked localization of tissue destruction in parts of the hippocampus. In contrast, infection with the Dearing strain caused tissue destruction not only in the hippocampus but also diffusely in the cortex and brain stem (unpublished data). Thus the pattern of restricted growth in the brain was not due simply to a general defect in the ability of the three variants to replicate, but to growth in restricted sites in the brain.

Thus, antigenic variants of reovirus type 3 selected by monoclonal antibodies directed against the major neutralization site, have markedly different levels of neurovirulence and capacity to grow *in vivo*. These experiments directly indicate that the reovirus type 3 haemagglutinin is a major determinant of neurovirulence, and imply that particular regions of the HA have critical roles in virulence. The site chosen for selection of the variants, the major neutralization determinant on the HA molecule, is probably the polypeptide region involved in binding to neurone receptors[14]. Other studies have shown that this region is also the site on HA that is recognized by cytotoxic T lymphocytes (unpublished data), and is involved in binding to cell receptors[14]. Furthermore, we have recently demonstrated that these variants induce cytolytic T lymphocytes as well as neutralizing antibody (unpublished data). Further data on the nature of the alterations in the haemagglutinin polypeptide of these variants should enable us to define the role of haemagglutinin in viral virulence more precisely. In addition, the use of monoclonal antibodies to generate attenuated viral strains may serve as a rapid and simple step for developing viral vaccines.

We thank Elaine Freimont for technical assistance and David Knipe and Robert Finberg for their helpful suggestions concerning the manuscript. This work was supported by grant 2 ROI AI 13178-06 from the National Institute of Allergy and Infectious Diseases, NIH. D.R.S. was supported by a fellowship from the Juvenile Diabetes Foundation.

Received 28 October 1981; accepted 19 March 1982.

1. Margolis, G. L., Kilham, L. & Gonatos, N. K. *Lab. Invest.* **24,** 101–109 (1971).
2. Raine, C. S. & Fields, B. N. *J. Neuropath. exp. Neurol.* **32,** 19–33 (1973).
3. Weiner, H. L., Drayna, D., Averill, D. R. & Fields, B. N. *Proc. natn. Acad. Sci. U.S.A.* **74,** 5744–5748 (1977).
4. Weiner, H. K., Powers, M. L. & Fields, B. N. *J. infect. Dis.* **141,** 609–618 (1980).
5. McCrae, M. A. & Joklik, W. K. *Virology* **89,** 578–593 (1978).
6. Mustoe, T. A., Ramig, R. F., Sharpe, A. H. & Fields, B. N. *Virology* **89,** 594–604 (1978).
7. Weiner, H. L. & Fields, B. N. *J. exp. Med.* **146,** 1303–1310 (1977).
8. Weiner, H. K., Ramig, R. F., Mustoe, T. A. & Fields, B. N. *Virology* **86,** 581–584 (1978).
9. Finberg, R., Weiner, H. L., Fields, B. N., Benacerraf, B. & Burakoff, S. J. *Proc. natn. Acad. Sci. U.S.A.* **79,** 442–446 (1979).
10. Finberg, R., Weiner, H. L., Burakoff, S. J. & Fields, B. N. *Infect. Immunity* **31,** 646–649 (1981).
11. Babiss, L. E. *et al. J. Virol.* **30,** 863–874 (1979).
12. Sharpe, A. H. & Fields, B. N. *J. Virol.* **38,** 389–392 (1981).
13. Burstin, S. J., Spriggs, D. R. & Fields, B. N. *Virology* **117,** 146–155 (1982).
14. Nepom, J. T. *et al. J. exp. Med.* **155,** 155–167 (1982).
15. Reed, L. J. & Muench, H. *Am. J. Hyg.* **27,** 493–497 (1938).
16. Ahmed, R. *et al. Cell* **25,** 325–332 (1981).

Non-neutralizing monoclonal antibodies can prevent lethal alphavirus encephalitis

A. L. Schmaljohn*, E. D. Johnson†, J. M. Dalrymple† & G. A. Cole*

* Department of Microbiology, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA
† Virology Division, US Army Medical Research Institute of Infectious Diseases, Fort Detrick, Frederick, Maryland 21701, USA

Among the heterogeneous population of antibodies specifically induced during many acute viral infections, those having virus-neutralizing activity *in vitro* are generally considered to be most important for recovery and immunity to reinfection. Similarly, the ability to stimulate production of circulating neutralizing (NT) antibodies is a major criterion for evaluating the immunoprophylactic potential of many antiviral vaccines[1]. Although there is obviously an association between NT antibody induction and host resistance, we present here data which indicate that other virus-specific antibodies lacking NT function may be equally important in conferring protective immunity to alphaviruses. We used monoclonal antibodies against Sindbis virus (SV) to demonstrate that passive protection of SV-infected mice from fatal paralytic central nervous system (CNS) disease may be mediated not only by antibodies which neutralize the infectivity of extracellular virus particles but also by those lacking this capacity, which react preferentially with virus-infected cells.

Sindbis virus, the prototype of the alphaviruses, contains two distinct envelope glycoproteins, E1 and E2, which exist as a complex, and a nucleocapsid protein, NC. It has been shown that antibodies raised in rabbits against each of the individually isolated and purified glycoproteins of SV possess distinctly separable activities *in vitro*[2]. Antibodies to E1 (the viral haemagglutinin) inhibit haemagglutination by SV, whereas only antibodies to E2 significantly neutralize viral infectivity. Antibodies specific for the internal NC do not react with intact virions. Both haemagglutination-inhibiting (HI) and NT antibodies are induced during productive infections with SV as well as with other alphaviruses[3], but their respective roles in mediating host recovery are not clear.

Recently[4], we described the preparation of a large panel of monoclonal antibodies against the parental AR339 strain of SV (SV_p), some of which were used to define serological and biochemical relationships between SV and certain other alphaviruses. In the present study, six anti-E1, six anti-E2, and one anti-NC monoclonal antibodies were selected for further characterization and each was assessed for its protective capacity *in vivo*. Because, by any route of inoculation, SV_p produces an abortive immunizing infection in adult mice, a neuroadapted (na) variant[5] of SV_p was used to intracerebrally (i.c.) challenge C57BL/6J (B6) mice previously given monoclonal antibodies. When given i.c. to non-immune adult B6 mice, SV_{na} produces paralysis and fatal encephalitis.

To prepare monoclonal antibodies, spleen cells from BALB/cJ mice immunized with infectious SV_p were fused with P3X63-Ag8 myeloma cells, and hybrid cells were cloned by limiting dilution in hypoxanthine–aminopterin–thymidine medium[6]. Clones producing antibodies to SV_p were selected by solid-phase radioimmunoassay[7] of supernatants. Positive cultures were subcloned in soft agar and tested again by radioimmunoassay for anti-SV activity. Individual hybrid clones were expanded *in vitro*, and then inoculated into groups of pristane-primed BALB/cJ mice[8]. The resulting ascitic fluids derived from each hybrid clone were pooled, frozen in aliquots, and used in subsequent experiments. The specificity of each monoclonal antibody was determined by radioimmunoassay reactivity with individual SV_p structural proteins which had been separated by isoelectric focusing as previously described[2]. Specificity was confirmed by immunoprecipitation of radiolabelled virus, followed by polyacrylamide gel electrophoresis.

The isotypes of the monoclonal antibodies were determined by an indirect immunofluorescence procedure in which glass slides bearing acetone-fixed, SV_p-infected BHK-21 cells were exposed sequentially to predetermined concentrations of monoclonal antibody, rabbit antiserum (Litton Bionetics) specific for individual mouse immunoglobulin heavy-chain isotypes (IgM, IgG1, IgG2a, IgG2b or IgG3), and fluorescein-conjugated goat

0028-0836/82/180070—03$01.00

Table 1 Properties of anti-SV_p monoclonal antibodies

No.	Monoclonal antibody Specificity	Isotype	Reactivity with SV_{na} IF	NT titre	% C′-dependent cytolysis
		Protective *in vivo*			
12	E2	IgG2b	+	3,200	55.3
43	E2	IgG2a	+	800	21.1
49	E2	IgG2a	+	1,000	47.7
1	E1	IgG2a	+	<10	29.4
2	E1	IgG2a	+	<10	16.8
7	E1	IgG2a	+	<10	29.1
38	E1	IgG2a	+	<10	32.6
31	E1	IgG2b	+	<10	28.5
		Non-protective *in vivo*			
23	E2	IgG3	+	<10	2.1
8	E2	IgG2a	−	<10	ND
10	E2	IgG2a	−	<10	ND
13	E1	IgG1	+	<10	−1.3
3	NC	IgG3	+	<10	0.3

Specific binding by monoclonal antibody to SV_{na}-infected BHK-21 cells was determined by immunofluorescence (IF). Acetone-fixed cells on glass slides were overlaid with monoclonal antibody for 30 min, washed and then overlaid with fluorescein-labelled goat anti-mouse immunoglobulins (Cappel) for 30 min, then washed again. Each reagent was used at a 1:20 dilution in buffered saline, *p*H 7.4. All positive monoclonal antibodies exhibited bright perinuclear and cytoplasmic IF; antibodies 8 and 10 were completely negative. No IF was observed for uninfected control cells. Using the same procedure, all monoclonal antibodies including 8 and 10, were positive on SV_p-infected cells. To demonstrate C′-dependent antibody-mediated cytolysis, BHK-21 cells were infected (or mock-infected) in suspension with virus at a multiplicity of 6 PFU per cell, labelled with ^{51}Cr, washed and seeded into flat-bottomed 96-well tissue culture plates (Costar). Each well received an aliquot (100 μl) containing 2×10^4 cells in RPMI-1640 supplemented with 10% fetal bovine serum. To triplicate wells were added monoclonal antibody (previously heated at 56 °C for 30 min) and guinea pig C′ at final concentrations (v/v) of 2.5% and 5.0%, respectively, in a total volume of 200 μl per well. After 6 h incubation at 37 °C in air containing 5% CO_2 (9 h post-infection) supernatants from each well (100 μl) were counted individually for radioactivity, and the mean % virus-specific cytolysis was calculated as described previously[15]; s.e.m. never exceeded 3%. Monoclonal antibody-mediated release of label from cells in control wells (uninfected cells + C′, and infected cells + heat-inactivated C′) did not exceed 2.8%. In medium alone, 17.9% and 16.7% of total incorporated label were released by infected and uninfected cells, respectively.

anti-rabbit immunoglobulins (Cappel). Slides were washed exhaustively after the application of each reagent.

Each monoclonal antibody was tested, in the presence of 5% (v/v) guinea pig complement (C′), for its ability to neutralize SV infectivity and to mediate C′-dependent lysis of SV-infected BHK-21 cells. We measured neutralizing activity against 100 plaque-forming units (PFU) of each virus preparation (SV_p and SV_{na}), using a conventional 'constant virus–varying antibody' assay[8]. The reciprocal of the highest dilution of a monoclonal antibody which produced a ≥80% reduction in PFU was designated its NT titre. SV-specific antibody-mediated cytolysis was determined in a 6-h ^{51}Cr-release assay (Table 1).

To initially screen monoclonal antibodies for their protective capacity, each was given intraperitoneally (0.1 ml per mouse) to five female B6 mice, 6–8 weeks old, 24 h before challenge i.c. with 1,000 PFU of SV_{na}. All, or most mice in eight such groups resisted challenge, whereas all of a control group of infected non-immune mice, died. As shown in Table 1 (upper panel), passive protection by three anti-E2 monoclonal antibodies correlated with their having measurable NT activity against SV_{na}, although all such antibodies were capable of neutralizing SV_p (data not shown). In contrast, although none of the six anti-E1 monoclonal antibodies had NT activity against either SV_p or SV_{na}, five were protective. Sindbis virions acquire their envelopes by budding from host-cell plasma membranes which express both E1 and E2 during viral morphogenesis[9,10]. Thus, irrespective of their glycoprotein specificity, all protective monoclonal antibodies shared an ability to bind SV_{na}-infected cells as shown by immunofluorescence and, as might be expected from their IgG2a and IgG2b isotypes, by C′-dependent cytolysis (Table 1).

The results of these initial passive protection tests were confirmed using larger numbers of B6 mice; protective monoclonal antibodies either prevented or delayed both paralytic disease and death (Table 2, upper panel). Measurements of infectious SV_{na} in individual brains obtained at intervals after challenge revealed that CNS infections were generally attenuated, rather than prevented, in mice which had received protective monoclonal antibody. On day 3 of infection, when brain virus titres were maximal in mice from groups given either non-protective or no monoclonal antibody, brains of mice from the protected groups contained at least 100-fold less virus (Table 2, lower panel). Furthermore, by day 7 of infection, when all unprotected mice were paralysed and/or moribund, infectious virus was no longer detectable in brains of protected mice (data not shown).

All monoclonal antibodies, whether protective or not, had similar titres ($>10^4$) against SV_p in solid-phase binding assays (data not shown), and the failure of a monoclonal antibody to protect *in vivo* could not be attributed, in any case, to transfer of insufficient antibody. Certain monoclonal antibodies may be nonprotective, however, due to antigenic differences between SV_p and SV_{na} that result in undetectable or low-affinity binding to SV_{na}. Of the three anti-E2 monoclonal antibodies which failed to protect the mice, none neutralized SV_{na} and only one (no. 23) was reactive with SV_{na}-infected cells by immunofluorescence (Table 1, lower panel). The non-C′-fixing isotype (IgG3) of the latter would explain its failure to mediate SV-specific cytolysis *in vitro* but the fact that all three neutralized SV_p infectivity, even in the absence of complement (data not shown), may mean that passive protection conferred by anti-E2 antibodies depended more on their having NT activity against SV_{na} and less on their isotypes. On the other hand, it seems significant that the isotype (IgG1) of the single non-protective anti-E1 monoconal antibody (no. 13) is one which does not mediate immunospecific cytolysis by the classical complement pathway[10]. If, in fact, protection by anti-E1 antibodies, which do not neutralize SV, is effected solely through this C′-dependent cytolytic mechanism, it would be restricted to only those isotypes capable of fixing complement. The importance of the latter in antibody-mediated protection against SV is also suggested by the impaired virus clearance previously noted in mice which are deficient in either the C3 (ref. 11) or C5 (ref. 12) complement components.

The present data suggest two specific but functionally different humoral mechanisms of SV elimination *in vivo*: (1) neutralization of infectivity by antibodies which bind to E2 determinants on extracellular virions, a process which can occur *in vitro* in the absence of C′; and (2) C′-dependent lysis of

Table 2 Prevention of paralysis and death of SV_{na}-infected mice by anti-SV_p monoconal antibodies

Monoclonal antibody No.	Speci-ficity	Paralysis P/T	Mean day of onset	Mortality D/T	Mean day of death	log_{10} PFU per g brain on day 3
12	E2	0/10	—	0/10	—	ND
43	E2	0/15	—	0/15	—	ND
49	E2	3/20	14	1/20	12	<2.0, 4.3, 2.7
1	E1	3/15	11	1/15	18	ND
2	E1	0/14	—	0/14	—	5.2, 4.6, 3.2
7	E1	0/14	—	0/14	—	5.2, 6.5, 5.5
38	E1	0/25	—	0/25	—	<2.0, <2.0, 2.8
31	E1	7/13	7	2/13	9	ND
23	E2	10/10	5	9/10	9	ND
8	E2	15/15	5	14/15	8	7.4, 7.7
10	E2	10/10	5	10/10	9	ND
13	E1	14/14	6	13/14	8	ND
3	NC	15/15	5	15/15	10	7.4, 7.5
None	—	19/20	6	18/20	9	7.7, 7.6

Groups of 6–8-week-old female C57BL/6J mice were given 0.1 ml monoclonal antibody (in 0.5 ml saline) intraperitoneally, challenged i.c. 24 h later with 1,000 PFU of SV_{na} and then observed for at least 21 days. Results are expressed as the proportion of mice that developed hind limb paralysis (paralysed/total, P/T) and died (dead/total, D/T). Brains of mice in selected groups were frozen 1, 3, 5 and 7 days after infection and assayed individually for SV PFU on BHK-21 cells; results are shown for day 3, when maximum brain titres occurred in non-immune control mice.

SV-infected cells mediated by either neutralizing or non-neutralizing antibodies (of restricted isotypes) which bind, respectively, to E2 and E1 determinants on the cell surfaces. It is likely that one or both of these mechanisms involve(s) the recognition of more than one viral glycoprotein determinant. For example, all six anti-E1 monoclonal antibody were capable of binding to SV-infected cells (Table 1) but only two (1 and 31) had significant HI activity (data not shown).

The ability of five anti-E1 monoclonal antibodies to effect cytolysis, but not neutralization, in the presence of C′ suggests that one or more E1 determinants expressed on the plasma membrane of infected cells are not similarly expressed on infectious SV particles. That non-NT anti-E1 antibodies did not recognize infectious SV was indicated further by their failure to neutralize virus in the presence of polyvalent anti-mouse-immunoglobulin antiserum which, in parallel assays, augmented the activity of NT antibodies (data not shown).

During viral morphogenesis, glycoproteins of SV and other alphaviruses undergo configurational changes associated with the cleavage of PE2 (the precursor of E2) and the subsequent formation of the virion E1/E2 complex[9,13]. These changes may cause certain E1 and/or E2 determinants expressed initially on infected cells to become modified or cryptic on mature virions. Therefore, it would follow that the determinant specificities of antibodies induced by a given alphavirus glycoprotein would depend, in part, on the protein being either cell- or virion-associated when presented to the immune system. In this context, our finding that the protective non-neutralizing anti-E1 monoclonal antibody was functionally specific for SV-infected cells suggests that the induction of such antibodies might be enhanced by immunization with infectious rather than inactivated virus.

Our results also bear on the long-standing phenomenon of cross-protective immunity between heterologous alphaviruses in the absence of antibodies having corresponding cross-reactive NT activity (for review see ref. 3). Although the basis of this phenomenon has been ascribed to cross-reactive virus-specific effector T cells[14], the participation of a humoral mechanism is suggested by the fact that monoclonal antibodies to the E1, but not the E2 glycoprotein, of SV exhibit serological cross-reactivity with other alphaviruses[4]. Whether these same non-neutralizing, monoclonal antibodies also mediate heterologous protection *in vivo* is a subject for future investigation.

We thank M. K. Gentry for preparing anti-SV monoclonal antibodies, Celia Jarvis for help in preparing the manuscript, and Dr J. F. Smith for his critical comments. This work was supported by the US Army Research and Development Command and USPHS grant NS 17741.

Note added in proof: We have recently found that four anti-SV monoclonal antibodies (including numbers 31 and 38 above), which bind three distinct determinants on E1, cross-react with western equine encephalitis virus and protect mice from lethal intraperitoneal challenge with this alphavirus.

Received 23 November 1981; accepted 5 March 1982.

1. Scherer, W. F. *et al. Am. J. trop. Med. Hyg.* **29**, 1359–1381 (1980).
2. Dalrymple, J. M., Schlesinger, S. & Russell, P. K. *Virology* **69**, 93–103 (1976).
3. Porterfield, J. S. in *The Togaviruses: Biology, Structure, Replication* (ed. Schlensinger, R. W.) 13–46 (Academic, New York, 1980).
4. Dalrymple, J., Cole, G., Jahrling, P., Johnson, E. & Harrison, S. *Abstr. 5th int. Congr. Virol.*, 380 (1981).
5. Griffin, D. E. & Johnson, R. I. *J. Immun.* **118**, 1070–1075 (1977).
6. Kennett, R. H., McKearn, T. J. & Bechtol, K. B. (eds) *Monoclonal Antibodies* 365–367 (Plenum, New York, 1980).
7. Zollinger, W. D., Dalrymple, J. M. & Artenstein, M. S. *J. Immun.* **117**, 1788–1798 (1976).
8. Monath, T. P. in *Manual of Clinical Immunology* (eds Rose, N. R. & Friedman, H.) 456–462 (American Society for Microbiology, Washington, 1976).
9. Scheefers, H., Scheefers-Borchel, U., Edwards, J. & Brown, D. T. *Proc. natn. Acad. Sci. U.S.A.* **77**, 7277–7281 (1980).
10. Spiegelberg, H. L. *Adv. Immun.* **19**, 259–294 (1974).
11. Hirsch, R. L., Griffin, D. E. & Winkelstein, J. A. *J. Immun.* **121**, 1276–1278 (1978).
12. Hirsch, R. L., Griffin, D. E. & Winkelstein, J. A. *Infect. Immunity* **30**, 899–901 (1980).
13. Schlesinger, M. J. & Kääriäinen, L. in *The Togaviruses: Biology, Structure, Replication* (ed. Schlesinger, R. W.) 371–392 (Academic, New York, 1980).
14. Peck, R., Brown, A. & Wust, C. J. *J. Immun.* **123**, 1763–1766 (1979).
15. Hapel. A. J., Bablanian, R. & Cole, G. A. *J. Immun.* **124**, 1990–1996 (1980).

Specific opiate receptors on isolated mammalian gastric smooth muscle cells

K. N. Bitar & G. M. Makhlouf

Division of Gastroenterology and Departments of Medicine and Physiology, Medical College of Virginia, Richmond, Virginia 23298, USA

The predominant opioid peptides of the gut, Met-enkephalin and Leu-enkephalin are confined to neurones of the myenteric plexus[1–4]. Axonal projections from these neurones innervate the circular muscle layer running parallel to the smooth muscle cells[5]. An opioid tridecapeptide (dynorphin), first isolated from the porcine pituitary[6], is found in neurones of the submucosal plexus in the guinea pig; axonal projections from these neurones also innervate the circular muscle layer[7,8]. Dynorphin consists of an N-terminal Leu-enkephalin sequence coupled to a C-terminal octapeptide by an Arg–Arg sequence. Neural receptors for all three opioid peptides and for opiate narcotics have been identified in the myenteric plexus where they cause hyperpolarization of neurones[9,10], inhibition of acetylcholine (ACh) release[11] and blockade of muscle contraction induced by field stimulation[12–16]. Measurements of motor activity in circular muscle where enkephalinergic fibres are most abundant, suggest that opiates also have a direct contractile action[17–20]. We have tested this possibility in isolated gastric smooth muscle cells devoid of neural elements[21,22], and report here the presence of specific high-affinity opiate receptors on such cells of the guinea pig. The receptors mediate contraction as measured by image-splitting micrometry, and exhibit a rank order of potency and absolute values similar to those reported for neural receptors of the myenteric plexus. In view of the distribution of enkephalinergic fibres predominantly to gut smooth muscle, our results strongly support a direct neuro- transmitter role for enkephalins in this region.

Smooth muscle cells were isolated from the stomach of the guinea pig as described in detail elsewhere[23] (see Fig. 1 legend). The morphometric profile of freshly dispersed cells is illustrated in Fig. 1 (mean cell length ± s.e.m. = 109.3 ± 1.2 μm; $n = 52$). Exposure of the cells to maximally effective concentrations of Met-enkephalin, Leu-enkephalin or morphine for 30 s shifted the curve to the left, causing the appearance of a greater proportion of shorter, contracted cells. The contractile response to a given dose of agonist was determined from the fractional (percentage) or absolute decrease in mean cell length compared with controls.

The kinetics of contractile response to Met-enkephalin and other opiates were similar to those reported for other contractile agents such as ACh, COOH-terminal octapeptide of cholecystokinin (CCK-OP) and gastrin-17)[22]. The response was immediate, reached a peak in 30 s and reverted to a sustained plateau for at least 8 min (Fig. 2). Figure 3 shows dose–response curves constructed for opioid peptides and morphine using the values

Table 1 Maximum contractile responses to opiate agonists

	Maximal contractile response (% decrease in cell length)	Potency [D_{50} (M)]
Dynorphin	33.7 ± 1.9 ($n = 7$)	10^{-10}
Met-enkephalin	34.1 ± 1.4 (9)	1.2×10^{-9}
D-ala^2-met-enkephalinamide	30.1 ± 0.9 (5)	6.0×10^{-9}
Morphine	29.7 ± 1.3 (5)	2.0×10^{-8}
Leu-enkephalin	30.0 ± 3.6 (7)	1.2×10^{-7}

0028–0836/82/180072—03$01.00

of the 30-s peak responses. At the highest doses used, all opiate agonists elicited maximal contractile responses similar to those previously reported for 10^{-9} M CCK-OP and 10^{-7} M ACh[21,22] (30–34% decrease in mean cell length) (Table 1). The order of potency as determined from the dose required to produce one-half of the maximal response (D_{50}) was dynorphin > Met-enkephalin > D-Ala^2-Met-enkephalinamide > morphine > Leu-enkephalin. Note that the widest difference in potency (1,200-fold) was between Leu-enkephalin and dynorphin, which includes the sequence of Leu-enkephalin in its NH_2-terminal; this raises the possibility that tryptic cleavage of dynorphin to Leu-enkephalin terminates the action of dynorphin as a putative neurotransmitter. The higher potency of Met-enkephalin relative to its synthetic analogue suggests minimal or no degradation of Met-enkephalin in muscle cell suspensions[23]. If this assumption is correct, the potency as determined in isolated muscle cells should reflect the true

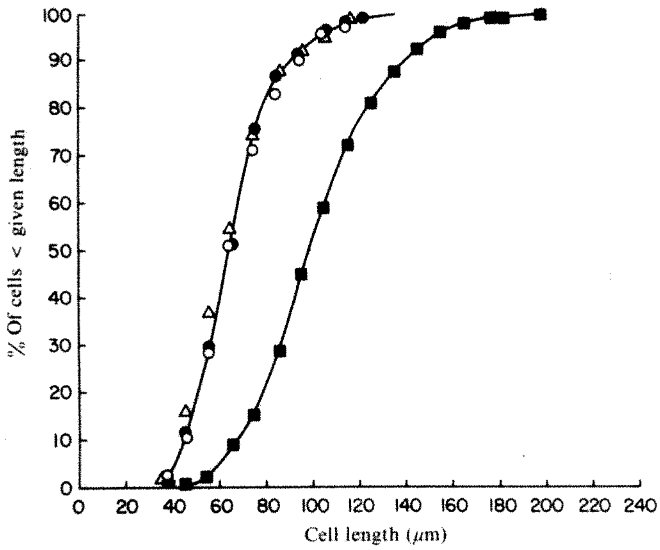


Fig. 1 Distribution of smooth muscle cell lengths in the control state (■) and after 30 s exposure to Met-enkephalin (●), Leu-enkephalin (△) or morphine (○). The shift of the curve to the left indicates the appearance of a greater proportion of shorter, more contracted cells. Cells were obtained from the circular muscle layer of the stomach of guinea pigs by two sequential 45 min incubations of muscle strips in Krebs bicarbonate buffer containing 0.1% collagenase (Worthington CLS, Type II) and 0.01% soybean trypsin inhibitor. The partly digested strips were washed with 50 ml of enzyme-free buffer and re-incubated for 30 min to allow cells to disperse spontaneously. The cells were collected by filtration through 500 μm Nitex then examined within 30 min of dispersion. Aliquots (2.5×10^4 cells in 0.5 ml) were incubated for 30 s with 0.1 ml of a solution containing agonist, and the cells were then fixed rapidly with acrolein. Incubation periods in the range 15 s to 8 min were used to examine the kinetics of contraction. Control cells were treated in the same manner but without added agonist. In the control state and for each test modality, the lengths of 50 cells in sequential microscopic fields were measured by image-splitting micrometry. Contraction was determined from the percentage decrease in mean cell length compared with control.

structure–activity status of Met-enkephalin and other opioid peptides.

Naloxone (3×10^{-10} M) caused a parallel rightward shift of the dose–response curve for Met-enkephalin without affecting maximal response (Fig. 4). A higher dose of naloxone (3×10^{-9} M) caused a further shift of the dose–response curve and 10^{-7} M abolished the response to all doses of Met-enkephalin. We compared the effect of 3×10^{-10} M naloxone against the highest effective doses of opiate agonists (10^{-6} M for dynorphin and Met-enkephalin; 10^{-5} M for morphine and Leu-enkephalin). Naloxone was a more effective inhibitor of the responses to morphine and Leu-enkephalin (83–85% inhibition) than of the responses to dynorphin and Met-enkephalin (34 and 24% inhibition, respectively). Naloxone inhibited

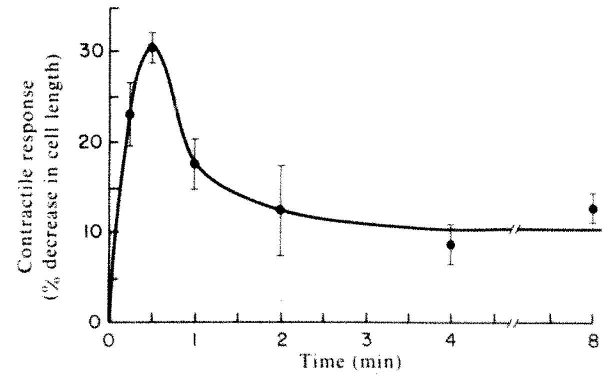


Fig. 2 Kinetics of contraction in response to 10^{-5} M Met-enkephalin. The prompt rise to a peak in 30 s and reversal to a sustained contractile plateau was similar to that previously found for non-opiate agonists[21,22]. Each point is the mean ± s.e.m. of four experiments.

specifically the response to opiate agonists; it affected neither the response to CCK-OP nor the response to ACh. Conversely, dibutyryl cyclic GMP (10^{-3} M), a specific antagonist of the CCK–gastrin receptor in neural, muscular and epithelial tissues[24,25], inhibited the contractile response to CCK-OP but had no effect on the response to the two opiate agonists tested (Met-enkephalin and morphine) or on the response to ACh[22]. Atropine (5×10^{-10}–5×10^{-8} M) also had no effect on the response to 10^{-6} M Met-enkephalin or morphine, although it abolished the response to ACh[22].

There was a remarkable similarity not only in the rank order but also in the absolute potencies of opiate agonists acting on isolated smooth muscle cells and nerve fibres of the guinea pig myenteric plexus[8,13,26,27]. In both tissues, dynorphin exhibited the highest potency ($D_{50} \sim 10^{-10}$ M) and Leu-enkephalin and morphine, the lowest potency ($D_{50} \sim 10^{-7}$ M). The potency of Met-enkephalin in isolated muscle cells was much closer to that of dynorphin than to that of Leu-enkephalin. The high potencies of dynorphin and Met-enkephalin and the relative resistance

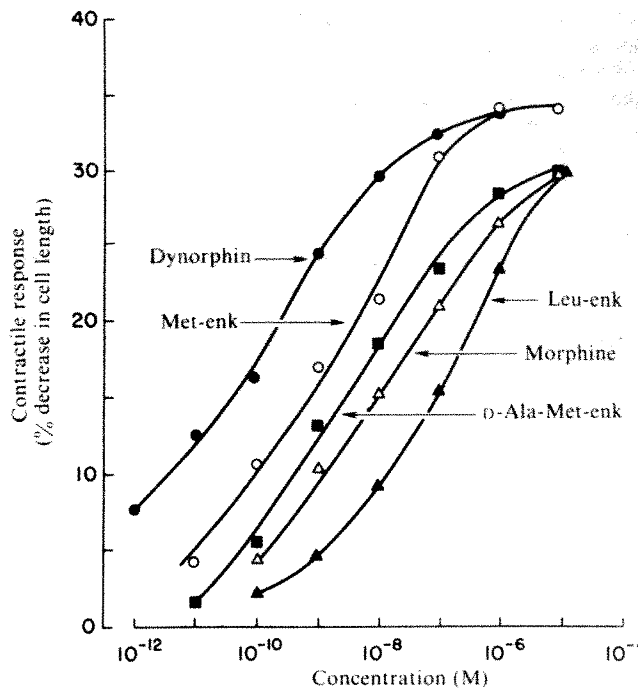


Fig. 3 Dose–response curves for the contractile effects of opioid peptides and morphine on isolated smooth muscle cells. Maximal responses and D_{50}s are given in Table 1. Each point is the mean of 4–9 experiments.

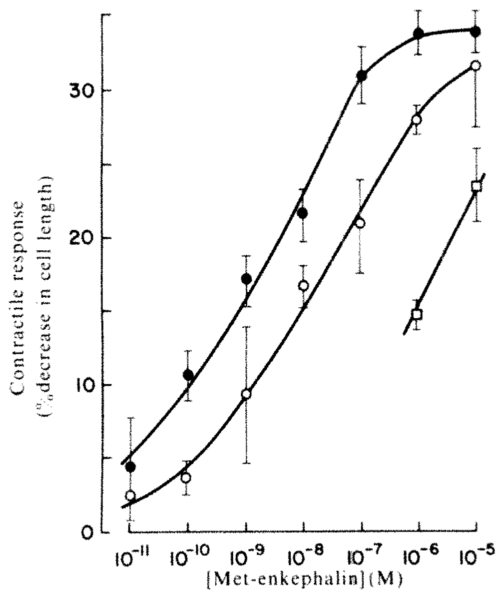


Fig. 4 Dose–response curves for the contractile effects of Met-enkephalin alone (●) or in combination with naloxone (○, 3×10^{-10} M; □, 10^{-9} M naloxone). Each point is the mean ± s.e.m. of four experiments.

of their effects to naloxone in isolated muscle cells suggest that the receptor(s) with which they interact are different from those for morphine (μ receptors). Chavkin and Goldstein[26] applied selective protection against the irreversible antagonist, chlornatrexamine, to show that the receptor for dynorphin in the myenteric plexus of the guinea pig is distinct from that for morphine or Leu-enkephalin. Concurrent desensitization of δ and μ receptors in mouse vas deferens also showed that the dynorphin receptor is distinct from that for other opiates[16,28].

The effectiveness of dynorphin and Met-enkephalin as direct contractile agonists agrees with their relative abundance in the stomach and intestine and their preponderance in the circular muscle layer. Similar results were obtained recently in smooth muscle cells isolated from the human stomach and canine gall bladder (K.N.B. and G.M.M., unpublished observations). The finding of a direct effect favours a role for these peptides as contractile neurotransmitters. However, dynorphin and Met-enkephalin are equally effective neural opiate agonists in the gut. Enkephalinergic fibers ramify with other fibres in the circular muscle layer, forming axonal bundles that could well permit axo-axonal interactions such as presynaptic inhibition of ACh release from cholinergic nerve terminals. It is even possible that enkephalins and ACh co-exist in some nerve terminals of the myenteric plexus as they do in sacral preganglionic fibres in the viscera of the cat[29] and postganglionic cholinergic terminals in the adrenal medulla[30]. Co-release, like adjacent release of enkephalins, could inhibit further presynaptic release of ACh.

Physiologically, the outcome is probably a combination of direct excitatory and indirect neurally-mediated inhibitory effects, the former inducing tonic motor activity in the gut, the latter, phasic or peristaltic activity[31]. The topography of enkephalin fibres in the gut indicates that a myenteric plexus–circular muscle preparation is the more apt physiological model with which to test neurally-mediated effects of enkephalin.

This work was supported by NIH grant AM-15564 and USPHS grant AM-28300.

Received 19 October 1981; accepted 23 February 1982.

1. Smith, T. W., Hughes, H., Kosterlitz, W. & Sosa, R. P. in *Opiates and Endogenous Opioid Peptides*, 57–62 (Elsevier/North-Holland, Amsterdam, 1976).
2. Furness, J. B., Costa, M., Franco, R. & Lewellyn-Smith, I. J. in *Neural Peptides and Neuronal Communication* (eds Costa E., & Trabucchi, M.) (Raven, New York, 1980).
3. Schultzberg, M. *et al. Neuroscience* **5**, 689–744 (1980).
4. Jessen, K. R. *et al. Neuroscience* **5**, 1717–1735 (1980).
5. Larsson, L. I., Childers, S. & Snyder, S. H. *Nature* **282**, 407–410 (1979).
6. Goldstein, A. & Ghazarossian, V. E. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6207–6210 (1980).
7. Watson, S. J., Akil, H., Ghazarossian, V. E. & Goldstein A. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1260–1263 (1981).
8. Goldstein, A., Tachibana, S., Lowney, L. I., Hunkapillar, M. & Hood, L. *Proc. natn. Acad. Sci. U.S.A.* **76**, 6666–6670 (1979).
9. Sakai, K. K., Hymson, D. L. & Shapiro, R. *Neurosci. Lett.* **10**, 317–322 (1978).
10. North, R. A. & Tonini, M. *Br. J. Pharmac.* **61**, 541–549 (1977).
11. Down, J. A. & Szerb, J. C. *Br. J. Pharmac.* **68**, 47–55 (1980).
12. Kosterlitz, H. W. in *Neural Peptides and Neuronal Communication* (eds Costa, E. & Trabucchi, M.) 633–642 (Raven, New York, 1980).
13. Lord, J. A. H., Waterfield, A. A., Hughes, J. & Kosterlitz, H. W. *Nature* **267**, 495–499 (1977).
14. Shaw, J. S. *Br. J. Pharmac.* **57**, 428–429 (1979).
15. Oka, T. & Tetsuo *Br. J. Pharmac.* **68**, 193–195 (1980).
16. Schulz, R., Wuster, M., Kerns, H. & Herz, A. *Nature* **285**, 242–243 (1980).
17. Nijkamp, R. P. & Van Ree, J. M. *Br. J. Pharmac.* **68**, 599–606 (1980).
18. Gillan, M. G. C. & Pollock, D. *B. J. Pharmac.* **68**, 381–392 (1980).
19. Scheurer, U., Drack, E., Varga, L. & Halter, F. *Regul. Peptides Suppl.* **1**, S100 (1980).
20. Burks, T. F. in *Physiology of the Gastrointestinal Tract* (ed. Johnson, L. R.) 495–516 (Raven, New York, 1981).
21. Bitar, K. N., Zfass, A. M. & Makhlouf, G. M. *Am. J. Physiol.* **237**, E172–E176 (1979).
22. Bitar, K. N. & Makhlouf, G. M. *Am. J. Physiol.* **240** (in the press).
23. Pert, C. B. & Pert, A. *Science* **194**, 330–332 (1976).
24. Peikin, S. R., Costenbader, C. L. & Gardner, J. D. *J. biol. Chem.* **254**, 5321–5327 (1979).
25. Hutchinson, J. B. & Dockray, G. J. *Regul. Peptides Suppl.* **1**, S53 (1980).
26. Chavkin, C. & Goldstein, A. *Nature* **291**, 591–593 (1980).
27. Kilpatrick, D. L. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 3265–3268 (1981).
28. Wuster, M., Schulz, R. & Herz, A. *Eur. J. Pharmac.* **62**, 235–236 (1980).
29. Glazer, E. & Basbaum, A. K. *Science* **208**, 1479–1481 (1980).
30. Di Gulio, A. M., Yang, H. Y., Fratta W. & Costa, E. *Nature* **278**, 646–647 (1978).
31. Kromer, W., Pretzlaff, W. & Woinoff, R. *Life Sci.* **26**, 1857–1865 (1980).

Opiate receptor subclasses differ in their conformational requirements

Peter W. Schiller & John DiMaio

Laboratory of Chemical Biology and Peptide Research, Clinical Research Institute of Montreal, 110 Pine Avenue West, Montreal, Quebec H2W 1R7, Canada

Pharmacological and biochemical studies of classical opiates and opioid peptide analogues have revealed the existence of several subclasses of opiate receptors[1–5], among which the μ-, δ-, κ- and σ-receptors have been most widely discussed. The physiological roles of the different receptor classes and their structural characteristics remain to be elucidated. Recently, a cyclic analogue of enkephalin, H-Tyr-cyclo[-N^{γ}-D-A_2bu-Gly-Phe-Leu-] (Fig. 1, compound I; A_2bu represents α,γ-diaminobutyric acid), showing high potency in μ-receptor-selective bio- and binding assays has been described[6]. To assess the effect of the conformational constraint introduced by cyclization on opiate receptor selectivity, we have compared the cyclic compound I with its corresponding open-chain analogue, [D-Abu^2,Leu^5]enkephalinamide (Fig. 1, compound II; Abu represents α-aminobutyric acid), in bioassays based on inhibition of electrically evoked contractions of the guinea pig ileum and the mouse vas deferens, and in binding assays using μ- and δ-receptor-selective radiolabels. The differences in potency of the two compounds observed in the four assay systems suggest that the various opiate receptor subclasses have different preferences in terms of conformational properties of 'complementary' ligands.

Compared with [Leu^5]enkephalin (H-Tyr-Gly-Gly-Phe-Leu-OH; III in Table 1)[7], the cyclic analogue I shows a 17-fold increase in potency in the guinea-pig ileum (GPI)-assay and is twice as potent as the linear analogue, II (Table 1). The effect of both analogues on GPI was found to be naloxone-reversible. In the case of compound I, the apparent dissociation constant (K_e) obtained with naloxone as antagonist was 12 times lower than that determined for [Leu^5]enkephalin and comparable to that observed for the μ-receptor-selective agonist, levorphanol. These observations indicate that the cyclic analogue exerts its effect on the GPI via interaction with μ-receptors.

In contrast to the results of the GPI assay, however, the cyclic analogue I is seven times less potent than the naturally occurring

0028–0836/82/180074—03$01.00

Table 1 Inhibitory potencies (IC_{50}s) and sensitivities to naloxone (K_e) of peptide analogues, [Leu[5]]enkephalin and levorphanol in the guinea pig ileum and mouse vas deferens assays

No.	Compound	IC_{50}(nM)* GPI	MVD	IC_{50} (MVD)/IC_{50} (GPI)	K_e (nM)* GPI	MVD
I	Tyr-cyclo[-N^{γ}-D-A_2bu-Gly-Phe-Leu-]	14.1±2.9	81.4±5.8	5.77	0.122±0.028	0.576±0.044
II	[D-Abu[2],Leu[5]]enkephalinamide	28.7±1.3	45.6±9.1	1.59	0.181±0.033	1.50±0.19
III	[Leu[5]]enkephalin	246±39	11.4±1.1	0.046	1.53±0.43	5.86±0.90
IV	Levorphanol	17.0±3.0	278±56	16.4	0.064±0.005	0.837±0.182

The assay based on inhibition of electrically induced contractions of the GPI[19] was performed as described elsewhere[20]. The MVD assay was performed as described previously[21]. A log dose–response curve was determined using [Leu[5]]enkephalin as the standard for each ileum or vas deferens preparation and IC_{50}s of the compounds being tested were normalized according to a published procedure[22]. K_e values for naloxone as antagonist were determined from the ratio (DR) of IC_{50} values obtained in the presence and absence of a fixed naloxone concentration (a), using the equation $K_e = a/(DR-1)$[23].

The chemical syntheses of analogues I and II are described elsewhere[6,17].

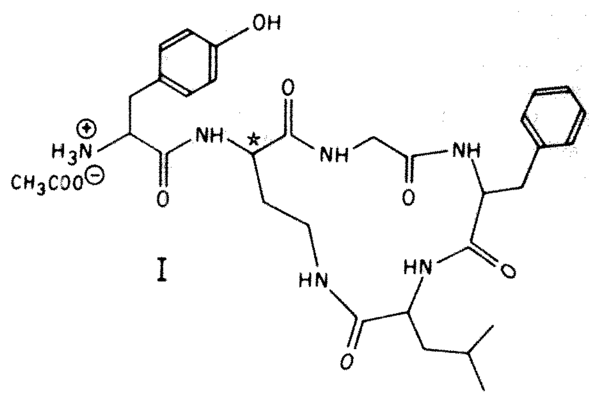
* Mean of three determinations (±s.e.m.).

peptide and half as potent as its corresponding open-chain analogue (II) in the mouse vas deferens (MVD)-assay. There is evidence that, in addition to the predominant δ-receptor, the MVD also contains μ- and κ-receptors. As indicated by the high naloxone K_e (5.86±0.90 nM), [Leu[5]]enkephalin interacts selectively with δ-receptors in the MVD preparation, whereas levorphanol, with a low K_e value of 0.837±0.182 nM, has high preference for μ-receptors in this tissue. The μ-receptor preference of the cyclic analogue I is revealed by a similarly low K_e (0.576±0.044 nM), whereas analogue II shows a K_e (1.50±0.19 nM) intermediate between those of levorphanol and [Leu[5]]enkephalin, indicating its mixed $\mu+\delta$ character.

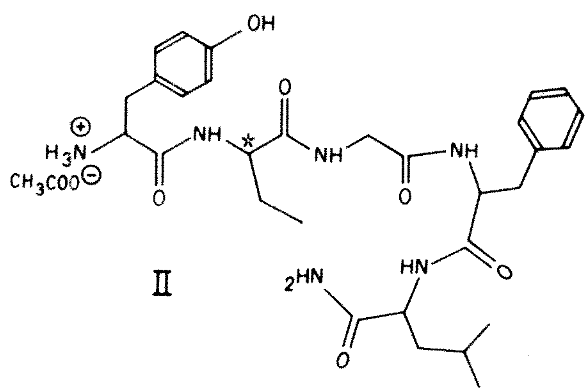
The MVD/GPI IC_{50} (50% inhibitory concentration) ratio observed for the cyclic analogue I was 125 times higher than that of [Leu[5]]enkephalin and was comparable to the ratios determined for the μ-type agonists, [D-Ala[2],MePhe[4],Met(O)-ol[5]]enkephalin[8] and [D-Ala[2],MePhe[4],Gly-ol[5]]enkephalin[9]. Interestingly, the cyclic compound I also showed a MVD/GPI IC_{50} ratio about four times higher than the open-chain analogue II. This result, together with the difference in the naloxone K_e values observed in the MVD assay, suggests that the conformationally restricted analogue I has a lower preference for δ-receptor than the more flexible open-chain analogue II.

Affinities for opiate receptors in rat brain membrane preparations were determined in parallel binding assays using ^{3}H-naloxone as a radioligand which preferentially binds to μ-receptors and ^{3}H-[D-Ala[2],D-Leu[5]]enkephalin as a radiolabel with preference for δ-receptors (Table 2). Sodium ratios of 47.0 and 38.9 were determined for analogues I and II, respectively, from ^{3}H-naloxone displacement curves obtained in the absence and presence (100 mM) of sodium chloride. This observation indicates the agonist character of both compounds[10]. In the δ-receptor-selective binding assay, analogue I showed a 28-fold decrease in affinity compared with analogue II, whereas in the μ-receptor-selective binding assay, only a twofold reduction in affinity was observed. The nearly equal affinity of the linear peptide II for μ- and δ-receptors and the low affinity of the cyclic compound I for δ-receptors are reflected in the ratios of the inhibition constants (K_i^{δ}/K_i^{μ}) obtained from the two binding assays. We observed some discrepancies between the binding data and the results of the bioassays; analogue I was twice as potent as its linear counterpart II in the GPI assay, whereas in the ^{3}H-naloxone binding assay, almost the reverse was true. A divergent effect of the conformational constraint in I on 'efficacy' and affinity represents one possible explanation for this observation. Cyclization would reduce affinity for μ-receptors but lead to a more productive peptide–receptor complex. On the basis of the well-known correlation between sodium ratio and 'efficacy'[10], the higher sodium ratio observed for analogue I compared with II seems to support this hypothesis. Alternatively, a third type of receptor may be involved and the simple μ/δ-receptor concept may be insufficient to explain the present data. That the situation in rat brain may indeed be more complicated is indicated by the recent demonstration[5] of three major types of morphine and enkephalin receptors (μ_1, μ_2 and δ). Finally, it is interesting to note that cyclization produces only a twofold potency decrease in the MVD assay but a 28-fold affinity decrease in the δ-receptor-selective binding assay. This apparent discrepancy can be explained by the fact that the greatly reduced affinity of the cyclic analogue for δ-receptors is partly compensated by its ability to act as a highly productive μ-agonist in the MVD, as indicated by its low naloxone K_e.

Thus, comparison of analogues I and II in bioassays and binding assays reveals that introduction of the conformational constraint by ring closure in analogue I is directly and uniquely responsible for its decreased δ-character compared with the linear peptide II. Obviously, the semi-rigid 14-membered ring structure of I is highly compatible with the topography of the μ-receptor, but causes a decrease in affinity for δ-receptors compared with the open-chain analogue II. Analysis of the conformational possibilities of I revealed that several conformations proposed for native enkephalin in solution or when bound to its receptor cannot be adopted by the cyclic analogue[6].



I



II

Fig. 1 Structural formulas of the cyclic enkephalin analogue, Tyr-cyclo[-N^{γ}-D-A_2bu-Gly-Phe-Leu-] (I), and its corresponding open-chain analogue, [D-Abu[2],Leu[5]]enkephalinamide (II).

Table 2 Inhibitory effects of peptide analogues and [Leu^5]enkephalin on the binding of ^{3}H-naloxone and ^{3}H-[D-Ala^2, D-Leu^5]enkephalin in rat brain homogenates

No.	Compound	^{3}H-naloxone binding, K_i^{μ} (nM)	^{3}H-[D-Ala^2,D-Leu^5]enkephalin binding, K_i^{δ} (nM)	K_i^{δ}/K_i^{μ}
I	Tyr-cyclo[-N^{γ}-D-A_2bu-Gly-Phe-Leu-]	13.8 ± 3.4	115 ± 16	8.33
II	[D-Abu^2,Leu^5]enkephalinamide	5.60 ± 0.11	4.06 ± 1.46	0.725
III	[Leu^5]enkephalin	27.7 ± 5.7	8.05 ± 0.96	0.291

Binding studies using rat brain membrane preparations were carried out as described elsewhere[20]. ^{3}H-naloxone and ^{3}H-[D-Ala^2,D-Leu^5]-enkephalin at concentrations of 0.40 and 0.96 nM, respectively, were used as radioligands and incubations were performed at 0 °C for 1 h. K_i values were calculated based on Cheng and Prusoff's equation[24] using values of 0.5 and 0.8 nM for the dissociation of ^{3}H-naloxone and ^{3}H-[D-Ala^2,D-Leu^5)enkephalin, respectively. Values shown are the mean of three determinations (±s.e.m.).

As the open-chain analogue is distinguished from the cyclic compound merely by the opening of a single carbon–nitrogen bond (Fig. 1), these results permit the unambiguous conclusion that different opiate receptor subclasses have different conformational requirements. The difference in receptor topography could be due to the existence of different conformational forms of a single opiate receptor protein, the distribution of which might be regulated by allosteric effectors[11]. Alternatively, heterogeneity in opiate receptor topography could be due to minor changes in receptor protein sequence[12].

The results described here represent the first direct demonstration of a difference in conformational requirements between receptor subclasses in a situation of peptide receptor heterogeneity. In the case of somatostatin[13,14] and bradykinin[15], differences in receptor topography between receptor types mediating different biological effects have been suspected but could not be demonstrated unambiguously, because in these studies, based on pharmacological profiles of peptide analogues, the effect of amino acid substitutions could not be dissociated from a possible conformational effect. Evidence for a difference in conformational requirements between different receptor subclasses has been obtained in the case of γ-aminobutyric acid[16].

Most linear peptides occurring in nature are very flexible and therefore able to adapt to different receptor surfaces without showing much selectivity. The demonstrated existence of receptor subclasses with topographically distinct features opens up an avenue for the design of receptor-selective peptide ligands by incorporation of appropriate conformational constraints. Subtle variation of the conformational restriction by modification of the ring size in analogue I[17] or by means of a different type of cyclization[18] has resulted in cyclic enkephalin analogues displaying various degrees of receptor selectivity.

We thank C. Lemieux and T. M.-D. Nguyen for technical assistance. This research was supported by an operating grant from the MRC of Canada (MT-5655).

Received 8 December 1981; accepted 16 March 1982.

1. Martin, W. R., Eades, C. G., Thompson, J. A., Huppler, R. E. & Gilbert, P. E. *J. Pharmac. exp. Ther.* **197**, 517–532 (1976).
2. Gilbert, P. E. & Martin, W. R. *J. Pharmac. exp. Ther.* **198**, 66–82 (1976).
3. Lord, J. A., Waterfield, A. A., Hughes, J. & Kosterlitz, H. W. *Nature* **267**, 495–499 (1977).
4. Chang, K.-J. & Cuatrecasas, P. *J. biol. Chem.* **254**, 2610–2618 (1979).
5. Wolozin, B. L. & Pasternak, G. W. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6181–6185 (1981).
6. DiMaio, J. & Schiller, P. W. *Proc. natn. Acad. Sci. U.S.A.* **77**, 7162–7166 (1980).
7. Hughes, J. *et al. Nature* **258**, 577–579 (1975).
8. Roemer, D. *et al. Nature* **268**, 547–549 (1977).
9. Handa, B. K. *et al. Eur. J. Pharmac.* **70**, 531–540 (1981).
10. Pert, C. B. & Snyder, S. H. *Molec. Pharmac.* **10**, 868–879 (1974).
11. Bowen, W. D., Gentleman, S., Herkenham, M. & Pert, C. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4818–4822 (1981).
12. Martin, W. R. in *The Bases of Addiction* (ed. Fishman, J.) 395–410 (Dahlem Konferenzen, Berlin, 1978).
13. Veber, D. F. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 2636–2640 (1978).
14. Meyers, C. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 577–579 (1980).
15. Chipens, G. I. *et al. Int. J. Peptide Protein Res.* **18**, 302–311 (1981).
16. Krogsgaard-Larsen, P., Johnston, G. A. R., Lodge, D. & Curtis, D. R. *Nature* **268**, 53–55 (1977).
17. DiMaio, J., Nguyen, T. M.-D., Lemieux, C. & Schiller, P. W. *J. med. Chem.* (submitted).
18. Schiller, P. W., Eggimann, B., DiMaio, J., Lemieux, C. & Nguyen, T. M.-D. *Biochem. biophys. Res. Commun.* **101**, 337–343 (1981).
19. Paton, W. D. M. *Br. J. Pharmac.* **12**, 119–127 (1957).
20. Schiller, P. W., Lipton, A., Horrobin, D. F. & Bodanszky, M. *Biochem. biophys. Res. Commun.* **85**, 1332–1338 (1978).
21. Hughes, J., Kosterlitz, H. W. & Leslie, F. M. *Br. J. Pharmac.* **55**, 541–546 (1975).
22. Waterfield, A. A., Leslie, F. M., Lord, J. A. H., Ling, N. & Kosterlitz, H. W. *Eur. J. Pharmac.* **58**, 11–18 (1979).
23. Kosterlitz, H. W. & Watt, A. J. *Br. J. Pharmac.* **33**, 266–276 (1968).
24. Cheng, Y. C. & Prusoff, W. H. *Biochem. Pharmac.* **22**, 3099–3108 (1973).

Dopamine modulates a Ca^{2+}-activated potassium conductance in mammalian hippocampal pyramidal cells

Larry S. Benardo* & David A. Prince

Department of Neurology, Stanford University School of Medicine, Stanford, California 94305, USA

Dopamine (DA) is a neurotransmitter in the mammalian central nervous system which has proven or potential importance in such neurological disorders as parkinsonism, Huntington's chorea and epilepsy. Most of the electrophysiological data concerning the actions of DA in the brain have been obtained from studies in the caudate nucleus where DA produces neuronal depolarization and increased spike discharge[1], slow depolarization with decreased spike discharge[2,3] and an increase in apparent input resistance[3] and hyperpolarization with reduced firing rate[3]. The mechanisms underlying these effects have not been examined. The evidence suggests that the hippocampus receives a dopaminergic projection[4–6] and that DA inhibits most hippocampal neurones[7,8]. We have studied the effects of DA on CA1 hippocampal pyramidal cells (HPCs) *in vitro* and report here that DA causes prolonged inhibition associated with hyperpolarization and increased conductance. These effects seem to derive from induction of a Ca^{2+}-activated K^+ conductance, and would make DA effective in modulating the high frequency firing and burst generation which occurs normally in some neurones[9,10], and pathologically in HPCs during epileptogenesis[11].

Hippocampal slices were prepared and maintained as previously described[12]. Intracellular recordings were obtained with K-acetate (4 M)-filled microelectrodes having resistances of 25–60 MΩ. Freshly prepared DA (10^{-6}–10^{-4} M) in standard Ringer's solution (*p*H 7.4)[12] was applied either by pressure ejection of droplets from a broken micropipette on to dendritic areas of CA1 HPCs[13] or via superfusion (10^{-5} M); these two methods produced identical results. Neurones having stable resting membrane potentials > -55 mV for longer than 5 min, and action potential amplitudes >70 mV were selected for study. Current was delivered through an active bridge circuit and bridge balance which was constantly monitored and adjusted. Control data were obtained with electrodes in the extracellular position so that records from micropipettes which showed polarization artefacts or rectification could be disregarded. Membrane potential (V_m) was determined by assuming the bath potential to be 0 mV.

Changes due to DA application were observed in 16 of 21 CA1 HPCs; in 13 of these, membrane properties were characterized in detail before and after drug application. The initial mean input resistance (R_N) of this group of cells was 38.9 ± 10.7

*Present address: Columbia University College of Physicians and Surgeons, 630 West 168th Street, New York, New York 10032, USA.

0028–0836/82/180076—04$01.00

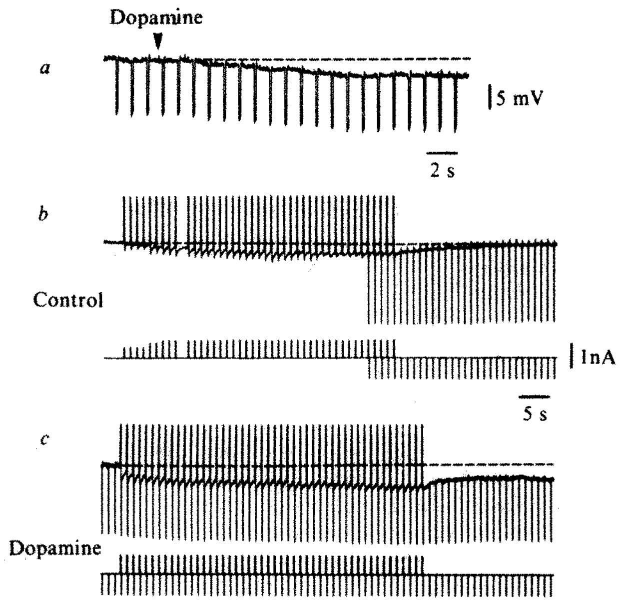


Fig. 1 Examples of DA-induced hyperpolarizations. *a*, Chart recording showing the hyperpolarizing response of a cell to application of 10 μM DA. The voltage response to constant-current hyperpolarizing pulses shows that no detectable change of input resistance (R_N) accompanies this hyperpolarization. Resting potential, −60 mV. *b*, In another neurone in normal conditions, a 1-Hz train of 120-ms depolarizing current pulses, of sufficient intensity to elicit 2–10 spikes, results in membrane hyperpolarization associated with a decrease in R_N. When depolarizing pulses are discontinued the membrane repolarizes and resistance recovers to the control level. *c*, After DA application to the cell in *b*, the membrane hyperpolarized by a few millivolts (data not shown). The membrane potential was then depolarized to the control level with d.c. current and a train of depolarizing pulses was delivered comparable to that in *b*. R_N decreased during the train. Membrane potential and resistance at the end of the train of depolarizing pulses did not recover to pre-stimulus levels. Control resting potential for the neurone in *b* and *c* was −56 mV. Time and current calibrations in *b* apply to *c* also. Voltage calibration in *a* applies to *b* and *c*. Spikes were amputated by inkwriter. Lower traces in *b*, *c*: current monitor, depolarizing current up. Broken lines represent baseline resting potentials.

(±s.d.) MΩ at a mean resting potential of -59.7 ± 5.7 mV. These values are similar to those previously reported[13]. DA application caused a small initial hyperpolarization (mean amplitude 4.9 ± 3.6 mV) in 10 of 13 cells within 3–40 s of delivery (Fig. 1*a*). Often this was not associated with a prominent change in R_N, although decreases in R_N were seen in cells which showed the largest hyperpolarizing responses, or when additional doses of DA were applied to produce further hyperpolarization. The shift in V_m was often sufficient to decrease firing frequency or silence active discharge. Further hyperpolarization was triggered by 120-ms depolarizing current pulses which evoked spike trains. It is known that slow after-hyperpolarizations (AHPs) lasting up to 1–2 s normally follow repetitive spike discharges in HPCs[9,10,14,15]. We found that, in control conditions, when depolarizing current pulses were delivered at frequencies of ~1 Hz, and at sufficient intensity to elicit 2–10 action potentials, these slow AHPs summated until steady membrane potential was reached (Fig. 1*b*). In the control state, V_m consistently recovered to the baseline level within seconds after such stimulation was discontinued. Following DA application, however, the slow AHPs evoked by the same depolarizing current pulses were larger and there was no recovery of V_m to the baseline potential after cessation of the pulses (Fig. 1*c*). In addition, slow AHPs following a single evoked spike train (tested at control V_m maintained with d.c. current) were also augmented (Fig. 2*a*–*c*). Such AHPs were of larger amplitude than those in control neurones and lasted for as long as 18 s (13 s in the case of Fig. 2*d*). Augmented AHPs were also found in neurones impaled after DA application (see below). The prolonged duration was not dependent on the passage of intracellular d.c. current. The DA effect on AHPs was detectable in seconds and reached a maximum after several minutes. In addition it appeared that the actions on the AHP were due to some cumulative process initiated by spike trains, as a sequence of conditioning trains evoked at 1 Hz by 120-ms depolarizing pulses increased the duration of an AHP triggered seconds later.

The maximum total hyperpolarization, that is, that which occurred after DA delivery (Fig. 1*a*) plus that which could be further evoked during summated AHPs (Fig. 1*c*), was 7.7 ± 6.0 mV ($n=13$). These peak increases in V_m were associated with decreases in R_N of 19–42% (compare amplitude of responses to hyperpolarizing pulses in Fig. 2*a*, *b*, before and after DA). The changes in R_N which accompanied the DA-induced hyperpolarization were not related to membrane rectification: for

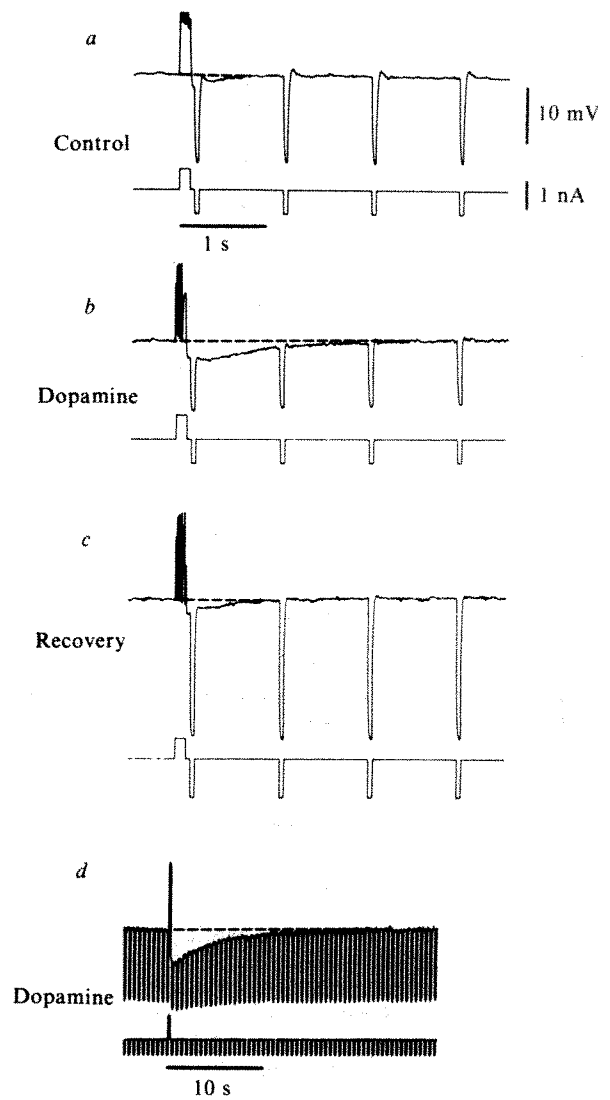


Fig. 2 Slow AHPs induced by depolarizing current pulses. *a*, Chart recording of control response to a depolarizing pulse. The AHP lasts 600 ms, reaches 2 mV in amplitude, and is associated with a peak decrease in resistance of ~11% (ref. 14). *b*, Following DA application (10 μM) this cell hyperpolarized (not shown) and input resistance decreased. The membrane was depolarized to the resting potential of *a* and the amplitude of the depolarizing current pulse adjusted to elicit a number of spikes comparable to those of the stimulus in *a*. This was necessary because of a decrease in baseline R_N of ~35% compared with the control. In these conditions AHPs lasted for 2.5 s, reached 4 mV in amplitude, and were associated with proportionately larger decreases in resistance (peak decrease of ~20%). The apparent difference in responses to initial depolarizing pulses in *a* compared with *b* is due to distortion produced by the low frequency response of the chart recorder, and the effect of DA in decreasing R_N. *c*, Recovery of AHPs to control duration and amplitude was seen in this same neurone 45–60 min after DA application. *d*, Much longer duration AHPs (>10 s) and associated resistance changes were evoked in several cells. In this neurone, impaled 1 h after DA application near the cell of *a*–*c*, the AHP lasted for 13 s. No d.c. current was passed through the electrode during this recording. Voltage and current calibrations in *a* apply to *b*–*d* also. Time calibration in *a* applies to *b* and *c*.

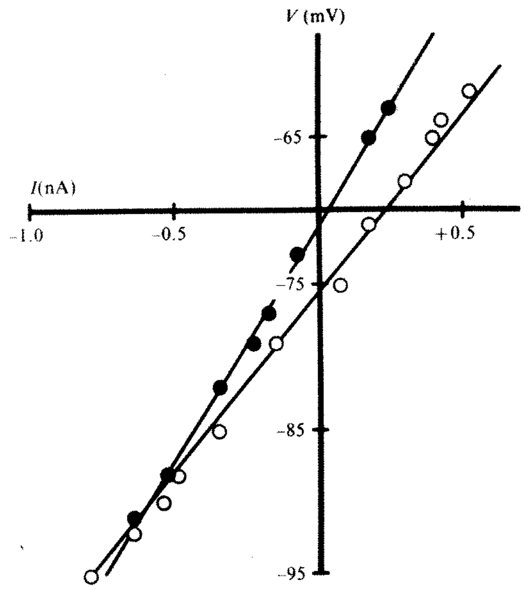


Fig. 3 Representative current–voltage relationship in one cell before (●) and after (○) DA application (10 μM). Curves were obtained by plotting steady-state applied pipette current against the membrane potential. Current (I)–voltage (V) lines were drawn using linear regression to obtain a least-squares fit of the data. The reversal potential for DA hyperpolarization in this neurone, determined from the point of intersection of these curves, was −91 mV.

example, the mean R_N before DA application was 37.3 ± 10.2 MΩ at the original resting V_m of -60.1 ± 6.0 mV ($n = 11$). After peak DA-induced hyperpolarization was attained, sufficient d.c. current was delivered to depolarize HPCs back to the original resting V_m. The mean R_N measured at this voltage level was 29.0 ± 8.8 MΩ, which was a significant ($P < 0.005$) decrease of 22.2% from control values.

A reversal potential for the DA hyperpolarization was obtained from plots of voltage against applied current before and after DA application. Figure 3 shows a representative plot from one neurone. The mean reversal potential obtained from intersection of voltage–current lines was -86.8 ± 11.1 mV ($n = 8$), a value consistent with DA increasing K^+ conductance (g_K)[15], and more negative than the expected Cl^- equilibrium potential of ~ -70 mV (ref. 16). This conclusion was supported by the observation that DA-induced increases in V_m were equally prominent in neurones impaled with KCl-containing microelectrodes, even when inhibitory postsynaptic potentials (i.p.s.ps), known to be primarily Cl^--mediated[17,18], were reversed.

In most cases, recovery from DA-induced hyperpolarization or augmentation of AHPs did not occur during a single impalement. However, in one particularly stable recording, it was possible to follow the time course of DA influence over a period of 2 h (Fig. 2). During the control period, AHPs reached a peak amplitude of 2 mV, and lasted ~600 ms (Fig. 2*a*). Following focal application of DA and adjustment of V_m to the control level with d.c. current, the AHP peak amplitude increased to 4 mV and lasted 2.5 s (Fig. 2*b*); recovery occurred in 45 min (Fig. 2*c*). Other neurones in the same area impaled up to 1 h after DA application showed similar or even more pronounced effects (Fig. 2*d*). No d.c. current was passed to manipulate V_m in cells impaled after DA application since resting V_m before DA application was unknown. The AHP durations and amplitudes, however, were significantly larger and longer than those found in control neurones in these and previous experiments[14]. These data suggest that the DA-induced alterations are reversible, but long-lasting.

Our results are similar in some respects to those reported by Herrling[8] for anaesthetized cats. In both cases cell discharges were decreased or blocked, and in most cells a hyperpolarization developed. However, Herrling found an increase in resistance in four neurones tested, in contrast to our results. The reasons for this discrepancy are not clear, but could relate to the action of anaesthetic or the technical difficulties in studying partially depolarized neurones *in vivo*.

The most reasonable explanation for the DA-induced hyperpolarization and the augmentation of AHPs is that they both derive from increases in K^+ permeability. As noted above, the equilibrium potential for DA hyperpolarization is in the range expected for this ion (Fig. 3). Effects on AHPs could be due to enhancement of the Ca^{2+}-activated g_K which underlies these events in HPCs[14,15]. DA-enhanced and normal AHPs are associated with conductance increases, have reversal potentials at very negative levels, and may both be blocked by Mn^{2+}, a Ca^{2+} antagonist[14,26]. If the DA-enhanced slow AHP were due in part to Na^+–K^+ pump activation[19], we would have expected persistence of at least a component of this potential when trains of spikes were evoked in Mn^{2+} solutions. Although all slow AHPs were blocked in 3 mM Mn^{2+} solutions, hyperpolarizations associated with conductance increases similar to that shown in Fig. 1*a* were still observed after DA application to some neurones. These effects must be postsynaptic as this concentration of Mn^{2+} blocks synaptic transmission in hippocampal slices[14].

Effects of DA in inducing or enhancing a Ca^{2+}-dependent K^+ conductance could result from effects on processes which increase intracellular Ca^{2+} concentration ($[Ca^{2+}]_i$) as proposed for neocortical neurones following dinitrophenol injection[20] and for cockroach neurones after DA application[21]. For example, the DA-induced initial hyperpolarization could result from release of Ca^{2+} from intracellular stores, or decreased Ca^{2+} buffering. These actions, combined with tonic Ca^{2+} entry[22], would prolong and increase K^+ conductance. Alternatively, DA may enhance calcium entry into neurones—this is suggested by its reported effects in cardiac muscle fibres[23]; however the persistence of DA-induced hyperpolarization in Mn^{2+} solutions that block Ca^{2+} entry[26] is more consistent with a DA-mediated release of Ca^{2+} from intracellular stores. The cumulative hyperpolarization which follows repetitive stimulus-induced spike trains (for example, Fig. 1*c*) would thus derive from prolonged increases in $[Ca^{2+}]_i$ following Ca^{2+} influx, which in turn would activate a sustained change in K^+ conductance. Recent experiments showing that EGTA injection into neurones blocks DA-induced hyperpolarization and DA-enhanced AHPs further substantiate the involvement of Ca^{2+} in DA effects[26]

DA is a potent inhibitory agent which modulates the relationship between cell activity and the development of intrinsic inhibitory membrane properties in HPCs through effects on the Ca^{2+}-activated K^+ conductance. Its inhibitory action would be most effective during periods of Ca^{2+} entry associated with burst generation[10] and high frequency firing such as occur in CA1 HPCs following exposure to acetylcholine[13], and during epileptogenesis[11,15,24]. The possibility that DA modulates such burst generation and seizure susceptibility *in vivo* is supported by anatomical data demonstrating a more or less direct projection of the hippocampal complex to the cells of origin of the dopaminergic projection[25].

We thank Drs B. W. Connors, L. M. Masukawa, B. R. Ransom and S. H. Thompson for helpful discussion, Ms C. Joo for secretarial assistance, and Mr J. L. Kadis for technical help. This work was supported by NS 12151 and NS 06477 (D.A.P.), a NSF Graduate Fellowship (L.S.B.), and the Morris Research Fund.

Received 23 July 1981; accepted 2 March 1982.

1. Kitai, S. T., Sugimori, M. & Kocsis, J. D. *Expl Brain Res.* **24,** 351–363 (1976).
2. Bernardi, G., Marciani, M. G., Morcutti, C., Pavone, F. & Stanzione, P. *Neurosci. Lett.* **8,** 235–240 (1978).
3. Herrling, P. L. & Hull, C. D. *Brain Res.* **192,** 441–462 (1980).
4. Swanson, L. W. & Hartman, B. K. *J. comp. Neurol.* **163,** 467–505 (1975).
5. Bischoff, S., Scatton, B. & Korf, J. *Brain Res.* **165,** 161–165 (1979).
6. Scatton, B., Simon, H., Le Moal, M. & Bischoff, S. *Neurosci. Lett.* **18,** 125–131 (1980).
7. Biscoe, T. J. & Straughan, B. *J. Physiol., Lond.* **183,** 341–359 (1966).
8. Herrling, P. L. *Brain Res.* **212,** 331–343 (1981).
9. Kandel, E. R. & Spencer, W. A. *J. Neurophysiol.* **24,** 243–259 (1961).

10. Wong, R. K. S. & Prince, D. A. *Brain Res.* **159**, 385–390 (1978).
11. Schwartzkroin, P. A. & Prince, D. A. *Ann. Neurol.* **1**, 463–469 (1977).
12. Schwartzkroin, P. A. *Brain Res.* **128**, 53–68 (1975).
13. Benardo, L. S. & Prince, D. A. *Brain Res.* **211**, 227–234 (1981).
14. Hotson, J. R. & Prince, D. A. *J. Neurophysiol.* **43**, 409–419 (1980).
15. Alger, B. E. & Nicoll, R. A. *Science* **210**, 1122–1124 (1980).
16. Anderson, P., Dingledine, R., Gjerstad, L., Langmoen, I. A. & Mosfeldt Laursen, A. *J. Physiol., Lond.* **305**, 279–296 (1980).
17. Krnjevic, K. *Physiol. Rev.* **54**, 419–540 (1974).
18. Alger, B. E. & Nicoll, R. A. *Nature* **281**, 315–317 (1979).
19. Segal, M. *Brain Res.* **206**, 107–128 (1981).
20. Godfraind, J. M., Kawamura, H., Krnjevic, K. & Pumain, R. *J. Physiol., Lond.* **215**, 199–222 (1971).
21. Ginsborg, B. L., House, C. R. & Mitchell, M. R. *J. Physiol., Lond.* **303**, 325–335 (1980).
22. Hotson, J. R., Prince D. A. & Schwartzkroin, P. A. *J. Neurophysiol.* **42**, 889–895 (1979).
23. Gelles, J. M. & Aronson, R. S. *Circulation Res.* **40**, 561–566 (1977).
24. Schwartzkroin, P. A. & Prince, D. A. *Brain Res.* **147**, 117–130 (1978).
25. Swanson, L. W. in *Functions of the Septohippocampal System*, 25–48 (Elsevier, Amsterdam, 1978).
26. Benardo, L. S. & Prince, D. A. *J. Neurosci.* (in the press).

Lipid-soluble toxins thought to be specific for Na^+ channels block Ca^{2+} channels in neuronal cells

Georges Romey & Michel Lazdunski

Centre de Biochimie du CNRS, Université de Nice, Faculté des Sciences, Parc Valrose, 06034 Nice Cedex, France

Many toxic compounds alter specifically the functioning of the fast sodium conductance that is responsible for the generation of action potentials in neuronal and muscle cells. One important class of these compounds consists of the lipid-soluble toxins, such as veratridine and other ceveratrum alkaloids, batrachotoxin, aconitine, and the diterpenoid grayanotoxins. These molecules are thought to bind to a common receptor site on the Na^+ channel[1–5], thus inducing a membrane depolarization which can be suppressed by the addition of tetrodotoxin (TTX) or by the removal of Na^+ from the incubation medium[3]. We show here that this family of molecules is not specific for the Na^+ channel in neuroblastoma cells. Blockade of the voltage-dependent calcium channel is also observed, either in the toxin concentration range in which the molecules act on the Na^+ channel (veratridine, grayanotoxin) or at even lower concentration (batrachotoxin).

Voltage-clamp experiments on differentiated N1E 115 neuroblastoma cells have revealed the existence of two voltage-dependent transient inward currents: a fast Na^+ current that is TTX-inhibitable and a slow Ca^{2+} current showing a voltage- and time-dependent inactivation[6]. Similarly to Ca^{2+} channels described in other preparations, the Ca^{2+} channel of neuroblastoma cells has the following properties: it is permeable to Ba^{2+} and Sr^{2+} cations[6], is inhibited by Mn^{2+}, Co^{2+}, Ni^{2+} or La^{3+} cations[6], is unaffected by TTX[6], and is blocked by the so-called Ca^{2+} channel blockers, verapamil or D_{600}, but only at concentrations as high as 0.1 mM (G.R., unpublished results). Voltage-clamp experiments presented in Fig. 1 show that veratridine and batrachotoxin cause a progressive decrease and ultimately a complete block of the Ca^{2+} current. The following are the main differences between the effects of veratridine and batrachotoxin: (1) batrachotoxin blocks the Ca^{2+} channel at a much lower concentration than veratridine. This difference will be analysed in more detail below. (2) The decrease of Ca^{2+} current due to the action of veratridine (Fig. 1*a*) is accompanied by a slowing of the inactivation step, whereas the kinetics of the Ca^{2+} channel are unchanged in the presence of batrachotoxin (Fig. 1*b*). (3) Veratridine blockage of the Ca^{2+} channel is rapidly reversible on washing for a few minutes (Fig. 2) whereas washing for more than 1 h does not remove the effects of batrachotoxin.

The peak amplitude of transient Ca^{2+} currents before and after applications of 20 and 200 μM veratridine is plotted as a function of membrane potential in Fig. 3*a*. Veratridine blockade of Ca^{2+} channels is partial at 20 μM and essentially complete at 200 μM, and is accompanied by a shift of the current–voltage curve towards more negative membrane potentials. Figure 3*b* shows dose–response curves for the action of veratridine and batrachotoxin in suppressing the transient Ca^{2+} current. The half-maximum effects ($K_{0.5}$) of the toxins are observed at concentrations of 26 μM and 40 nM for veratridine and batrachotoxin, respectively. Cevadine and α-dihydrograyanotoxin II are active in the same concentration range as veratridine.

The concentrations of veratridine and α-dihydrograyanotoxin II required for half-maximal stimulation of $^{22}Na^+$ influx in the same neuroblastoma cells were 44 and 90 μM, respectively[7,8]. Thus, veratridine and α-dihydrograyanotoxin II act on both Ca^{2+} channels and Na^+ channels in the same range of concentration. In contrast, the $K_{0.5}$ value for the action of batrachotoxin on the neuroblastoma Na^+ channel is ~1 μM (ref. 8), which is about an order of magnitude larger than that found for the action of the toxin on Ca^{2+} channels.

Dose–response curves for the action of veratridine, cevadine, α-dihydrograyanotoxin II or batrachotoxin on Na^+ channels in neuroblastoma cells have been obtained in the presence of 1.8 mM external Ca^{2+}. However, higher Ca^{2+} concentrations strongly antagonize this effect of these lipid-soluble toxins[3,5] and it is largely suppressed at 25 mM Ca^{2+}, the concentration at which Ca^{2+} currents were recorded here. Therefore, at high external Ca^{2+} concentrations, veratridine, α-dihydrograyanotoxin II and batrachotoxin seem to be much more specific for Ca^{2+} channels than for Na^+ channels.

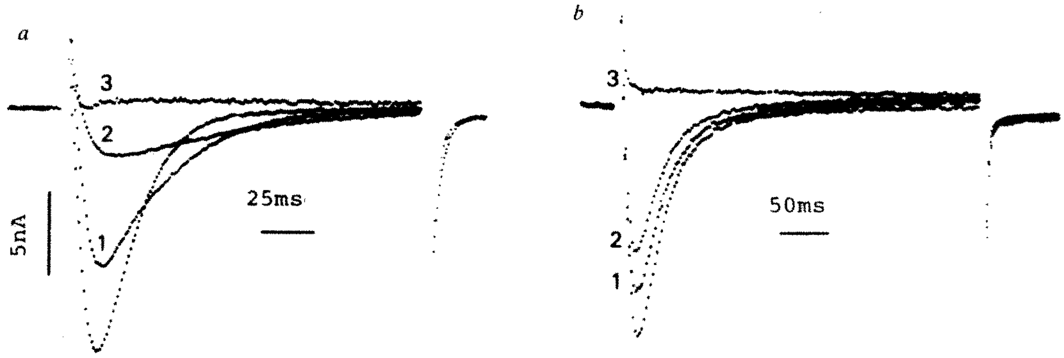


Fig. 1 Voltage-clamp analysis of the effects of veratridine (*a*) and batrachotoxin (*b*) on the Ca^{2+} current of N1E 115 neuroblastoma cells. Depolarizing pulses of 70 mV amplitude and 170 ms duration were applied from a holding potential of −80 mV. *a*, Time course of the effect of 2×10^{-4} M veratridine on the Ca^{2+} current. Successive recordings were taken 1 (1), 2 (2) and 5 (3) min after addition of veratridine. *b*, Time-course of the effect of 2×10^{-7} M batrachotoxin on the Ca^{2+} current. Successive recordings were taken 10 (1), 15 (2) and 30 (3) min after addition of batrachotoxin. Culture dishes containing N1E 115 neuroblastoma cells were used directly after the culture medium had been replaced by a Na^+-free solution containing 25 mM $CaCl_2$, 0.4 mM $MgSO_4$, 5.4 mM KCl, 25 mM HEPES-Tris, 5 mM glucose and 25 mM tetraethylammonium, adjusted to *p*H 7.4 and to an osmotic pressure of 305 mosmol with choline chloride. Tetraethylammonium was used to block the delayed outward K^+ current, which partially masked the Ca^{2+} current[6]. Voltage-clamp experiments were performed with the use of a suction pipette method analogous to that described by Lee *et al.* for snail neurones[10]. The intracellular perfusion solution was 10 mM NaH_2PO_4 and 115 mM glutamic acid, adjusted to *p*H 7.1 with KOH and to an osmotic pressure of 305 mosmol with sucrose.

0028-0836/82/180079—02$01.00

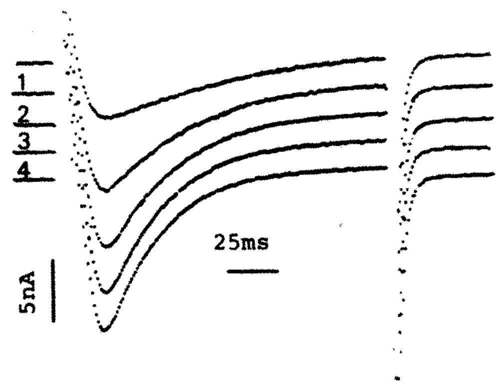


Fig. 2 Reversal by washing of the blocking effect of 10^{-4} M veratridine on the Ca^{2+} current. Successive recordings were taken 1 (1), 2 (2), 3 (3) and 4 (4) min after the beginning of washing. The recovered Ca^{2+} current in (4) was reduced by ~10% from the control Ca^{2+} current measured before the application of veratridine.

Aconitine and ceveratrum alkaloids of the veratridine family such as germitrine, germitridine, germine-3-acetate and protoveratrine A, which cause a persistent activation of Na^+ channels in neuroblastoma cells with an efficiency similar to that of veratridine (P. Honerjäger, C. Frelin and M. Lazdunski, in preparation), are inactive on Ca^{2+} channels, even at high concentrations (0.1 mM). Moreover, none of these molecules antagonizes the action of veratridine on the Ca^{2+} channel, so that they can be considered to be specific for the Na^+ channels.

Other toxins that act on the Na^+ channel include TTX and saxitoxin, and polypeptide toxins extracted from some scorpion venoms or sea anemones. TTX and saxitoxin block the increase of Na^+ permeability produced either by electrical stimulation or by the lipid-soluble toxins. Polypeptide toxins specifically slow the inactivation of the Na^+ channel without changing the properties of the activation step[4]. The polypeptide toxins act in synergy with the lipid-soluble toxins. In neuroblastoma cells they produce an increase in the fraction of Na^+ channels activated at a saturating concentration of the lipid-soluble toxins[7]. We have shown that high concentrations of TTX (10 μM), of *Androctonus australis* Hector scorpion toxin II (0.1 μM) and of *Anemonia sulcata* toxin II (10 μM), are without effect on Ca^{2+} channels in neuroblastoma cells. Moreover, the same high concentrations of these toxins had no effect on the blockade of the Ca^{2+} channels caused by veratridine, regardless of whether they were added before or after the lipid-soluble toxin.

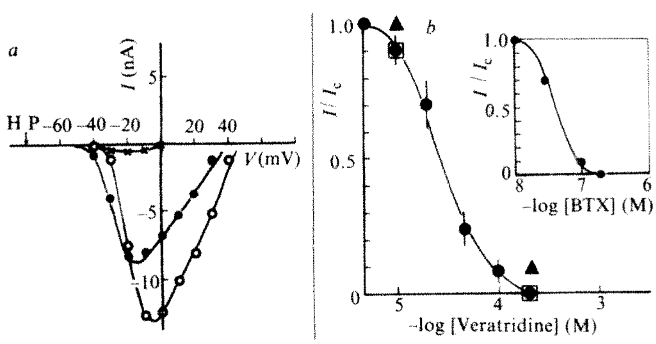


Fig. 3 *a*, Current-voltage relationships for the peak transient Ca^{2+} current obtained from the same cell before (○) and after a 10-min application of 2×10^{-5} M veratridine (●). Veratridine was then added at a final concentration of 2×10^{-4} M for 5 min (×). *b*, Effect of different concentrations of veratridine (●), cevadine (□), α-dihydrograyanotoxin II (▲) and batrachotoxin (inset) on the Ca^{2+} current. For each concentration of toxin, a depolarizing pulse of 70 mV was applied from a holding potential of −80 mV. I_C is the peak value of the Ca^{2+} current in the control, and I is the peak value in the presence of different amounts of toxin. For veratridine, each point represents the mean ratio of I/I_C (±s.d.) from five experiments.

The present observation that veratridine and batrachotoxin block Ca^{2+} channels may lead to a reinterpretation of some of the results obtained by various researchers who used these toxins as tools on the assumption that they were specific for the Na^+ channel.

The fact that veratridine, batrachotoxin, cevadine and α-dihydrograyanotoxin II act on both Na^+ and Ca^{2+} channels suggests that some common structural feature may be necessary for the operation of both types of channel. As tritiated batrachotoxin has now been synthesized[9] and as batrachotoxin is a very potent Ca^{2+} blocker, one of the most interesting possibilities for the future is that this toxin might become the first tool available for the biochemical titration of the Ca^{2+} channel.

Batrachotoxin was kindly provided by Dr J. Daly, α-dihydrograyanotoxin II by Drs M. Nakajima and Ryozo Irize and ceveratrum alkaloids by Dr P. Honerjäger. We thank Dr C. Frelin and M. T. Ravier for providing neuroblastoma cells and Dr Y. Tourneur for his collaboration. This work was supported by CNRS, DGRST and INSERM (CRL, 80.30.10).

Received 11 January; accepted 16 March 1982.

1. Narahashi, T. *Physiol. Rev.* **54**, 813–889 (1974).
2. Albuquerque, E. X. & Daly, J. W. in *The Specificity and Action of Animal, Bacterial and Plant Toxins* (ed. Cuatrecasas, P.) 299–338 (Chapman & Hall, London, 1976).
3. Ulbricht, W. *Ergebn. Physiol.* **61**, 18–71 (1969).
4. Lazdunski, M. *et al.* in *Neurotransmitters and their Receptors* (eds Littauer, U. Z., Dudai, Y., Silman, I., Teichberg, V. I. & Vogel, Z.) 511–529 (Wiley, New York, 1980).
5. Catterall, W. A. *J. biol. Chem.* **250**, 4053–4059 (1975).
6. Moolenaar, W. H. & Spector, I. *J. Physiol., Lond.* **292**, 307–323 (1979).
7. Jacques, Y., Fosset, M. & Lazdunski, M. *J. biol. Chem.* **253**, 7383–7392 (1978).
8. Jacques, Y., Romey, G., Cavey, M. T., Kartalovski, B. & Lazdunski, M. *Biochim. biophys. Acta* **600**, 882–897 (1980).
9. Brown, G. B. *et al. Cell molec. Neurobiol.* **1**, 19–40 (1981).
10. Lee, K. S., Akaike, N. & Brown, A. M. *J. gen. Physiol.* **71**, 489–508 (1978).

Expression of eukaryotic coding sequences in *Methylophilus methylotrophus*

John F. Hennam*, Angela E. Cunningham*, Geoffrey S. Sharpe† & Keith T. Atherton*

* Corporate Research Laboratory, ICI Ltd, Runcorn Heath, Cheshire WA7 4QE, UK
† ICI/University Joint Laboratory, Department of Biochemistry, University of Leicester, Leicester LE1 7RH, UK

Although there have been many reports of the expression of eukaryotic genes in *Escherichia coli* K-12 (refs 1–4), very little attention has focused on the expression of these sequences in other bacteria. The use of other organisms may prove advantageous, especially with regard to industrial processes. For example, work is in progress to develop alternative host/vector systems based on *Bacillus subtilis*[5] and *Saccharomyces cerevisiae*[6]. The obligate methylotroph, *Methylophilus methylotrophus*, is a fermentation organism which grows very well on cheap substrates, methanol and ammonia. Moreover, obligate methylotrophs are unlikely to infect man and hence may be considered relatively safe. Here we report the expression of two eukaryotic cDNAs encoding chicken ovalbumin and mouse dihydrofolate reductase in *M. methylotrophus*, hence demonstrating the potential of this organism for the commercial production of mammalian peptides.

Most cloning vectors developed to date are plasmids or phages specific for *E. coli*. Wide host-range cloning vectors based on RP4 and other IncP group plasmids have been developed[7], but these vectors are large, have a low copy number and are therefore not suitable for cloning when increased expression due to copy number is desired. For ease of manipulation and high gene dosage, small, high copy-number, broad

host-range plasmids are needed. IncQ group plasmids, including R300B[8], meet these criteria and although they are themselves non-conjugative, they are mobilized efficiently by conjugative plasmids of various groups[8]. Despite its advantages, however, R300B has relatively few restriction enzyme sites suitable for cloning. The composite plasmid, pGSS15[9], has both the ampicillin (Ap^r) and tetracycline (Tc^r) resistance from pBR322 while retaining the broad host-range characteristics of R300B and hence contains a number of unique restriction enzyme sites for cloning. Moreover, cloning of eukaryotic cDNA into the *Pst*I site of the Ap^r gene by dG.dC tailing has been demonstrated to lead to the expression of either fused[10] or native[11] eukaryotic polypeptides at least in *E. coli*.

It has been reported previously that using broad host-range cloning vectors, the *E. coli gdh* gene has been expressed in *M. methylotrophus*[12]. Here we have used pGSS15 to demonstrate expression of the mouse dihydrofolate reductase and chicken ovalbumin genes in *M. methylotrophus*.

Copy DNA derived from the mRNA for mouse dihydrofolate reductase (DHFR) has been cloned and expressed in *E. coli*[13]; bacteria which synthesize mouse DHFR can be selected because they are resistant to levels of trimethoprim that would ordinarily inhibit growth.

The plasmid pDHFR7 (ref. 13) was constructed by annealing deoxy-C-'tailed' cDNA to pBR322 DNA which had been cleaved in the β-lactamase gene by *Pst*I and 'tailed' with deoxy-G residues. The *Pst*I sites were regenerated on both ends of the insert as a result of such recircularization. The cDNA insert of pDHFR7 was isolated after *Pst*I cleavage of the plasmid and 1 μg of this DNA was mixed with 0.5 μg of pGSS15 DNA which had been linearized by *Pst*I digestion. The mixture was ligated and used to transform *E. coli* ED8654 (*met*B, *rk*$^-$*mk*$^+$, *sup*F). Out of 221 of the trimethoprim-resistant (Tp^r) clones, 216 were ampicillin-sensitive, showing that the DHFR fragment had been inserted into the *bla* gene. Gel electrophoretic analysis showed that all the plasmids in the Tp^{r}Ap^s clones were the same size and were larger than pGSS15. The DHFR cDNA has a single *Bgl*II restriction site which allows orientation of the fragment. *Eco*RI/*Bgl*II digests of plasmids from Tc^{r}Ap^{s}Tp^s and Tc^{r}Ap^{s}Tp^r clones established that only one orientation led to functional expression of the DHFR sequence, indicating that productive transcription was from the β-lactamase promoter.

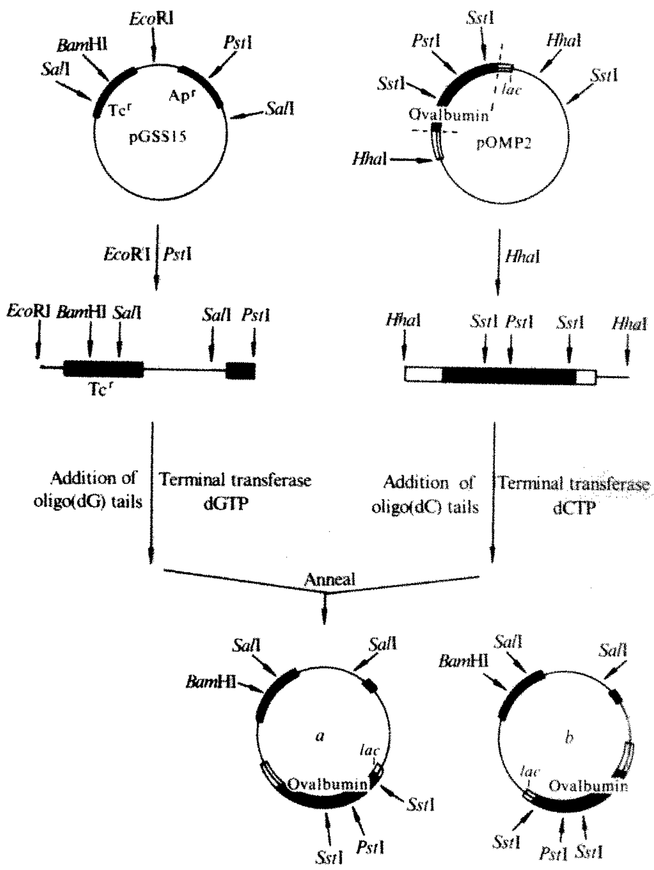


Fig. 1 Outline of the construction of pGSS15 derivatives carrying ovalbumin cDNA sequences fused to the *lac* UV5 promoter. pGSS15 DNA (1 μg) was digested with *Eco*RI and *Pst*I and the larger fragment isolated from an agarose gel using hydroxyapatite[25]. A *Hha*I DNA fragment containing the ovalbumin sequence and *lac* UV5 promoter from pOMP2 was isolated using the same technique. In the case of pOMP2, several other *Hha*I sites are present outside the *lac*–ovalbumin region (not shown). Oligo (dC) or (dG) tails were added using calf thymus terminal transferase, in the presence of 1 mM $CaCl_2$, and the two DNAs were mixed in equimolar amounts, at a total concentration of 1 μg ml^{-1}, for annealing[26]. Aliquots were used to transform $CaCl_2$-treated strain CL499 and transformants selected on nutrient plates containing (10 μg ml^{-1}) tetracycline. Colonies were tested for the presence of ovalbumin DNA sequence by hybridization to the *Taq*I ovalbumin fragment labelled with ^{32}P, and finally for the production of immunoreactive ovalbumin.

Table 1 Dihydrofolate reductase activities of pDHFR 2.43 in extracts of *E. coli* ED 8654 and *M. methylotrophus*

Strain	Substrate	Addition	Activity
E. coli ED 8654	Dihydrofolate	—	0.70±0.14 (4)
E. coli ED 8654 (pDHFR 2.43)	Dihydrofolate	—	1.58±0.26 (5)
M. methylotrophus	Dihydrofolate	—	2.34±0.68 (5)
M. methylotrophus	Dihydrofolate	Trimethoprim*	0 (2)
M. methylotrophus	Folate†	—	0 (2)
M. methylotrophus (pDHFR 2.43)	Dihydrofolate	—	12.6±1.7 (4)
M. methylotrophus (pDHFR 2.43)	Dihydrofolate	Trimethoprim*	2.57±0.53 (4)
M. methylotrophus (pDHFR 2.43)	Folate†	—	0.50 (1)

E. coli strains used were grown on medium containing (per l): $(NH_4)_2SO_4$, 2 g; K_2HPO_4, 14 g; KH_2PO_4, 6 g; trisodium citrate, 1 g; $MgSO_4.7H_2O$, 0.2 g (*p*H 7.2) supplemented with methionine (50 μg ml^{-1}) and glucose added to a final concentration of 0.2%. *M. methylotrophus* strains were grown on methanol minimal medium as described previously[12]. Cultures were grown with shaking at 37 °C overnight, collected by centrifugation and the pellets resuspended in 3 vol 50 mM potassium phosphate buffer (*p*H 7). The cells were disrupted by sonication on ice in the presence of 10 μM phenylmethylsulphonyl fluoride (PMSF) and 10 μM benzamidine and cell debris removed by centrifugation at 10,000*g* for 15 min and 100,000*g* for 60 min. Enzyme assays were performed at 37 °C. The standard assay mixture contained 50 μM dihydrofolate, 60 μM NADPH, 12 mM 2-mercaptoethanol, 50 mM potassium phosphate buffer *p*H 7 and enzyme (normally 100 μl of bacterial extract) in a final volume of 1.0 ml. The oxidation of NADPH was followed at 340 nm for ~10 min. Enzyme activity is expressed as nmol of NADPH oxidized per min per mg protein (mean ± s.e.m.). Numbers in parentheses are the number of separate experiments. Protein concentration was determined by the method of Lowry *et al.*[27].

* Trimethoprim (100 μg ml^{-1}) was added to the enzyme assays.

† Folate (70 μM) was included in the assay mixture instead of dihydrofolate.

The plasmid R64drd11 (ref. 14) was introduced into one Tp^r clone, pDHFR 2.43, and used to mobilize this hybrid plasmid into *M. methylotrophus*. Table 1 shows the levels of dihydrofolate reductase activity in extracts of *E. coli* (pDHFR 2.43) and *M. methylotrophus* (pDHFR 2.43). When trimethoprim (100 μg ml^{-1}) was added to the *M. methylotrophus* extracts, complete inhibition of the bacterial enzyme activity was observed, but the activity in the pDHFR 2.43 extract was reduced to only ~25%, consistent with observations that the mammalian enzyme is less prone to inhibition by trimethoprim[15]. Furthermore, changing the substrate to folate instead of dihydrofolate abolished the activity of the bacterial enzyme but the mammalian enzyme retained a low level of activity (Table 1 and ref. 16). These two observations confirm that the trimethoprim resistance of *M. methylotrophus* carrying pDHFR 2.43 is a consequence of the expression of the mouse dihydrofolate reductase gene.

As pGSS15 retains the broad host-range characteristics of R300B[8], we have been able to transfer pDHFR 2.43 into another methylotroph, *Methylomonas methylovora*, and into an autotrophic bacterium, *Alcaligenes eutrophus*, and have

Table 2 Synthesis of ovalbumin-like protein in bacterial extracts

Strain	Plasmid	Immunoreactive ovalbumin (molecules per cell)
E. coli CL499	pOV230	ND
E. coli CL499	pGSS15 : *Taq*-ov	2.8
M. methylotrophus	pGSS15 : *Taq*-ov	3.6
E. coli CL499	pOMP2	740
E. coli CL499	pGSS15 : *lac*-ov	760
M. methylotrophus	pGSS15 : *lac*-ov	150

Cultures (5 ml) of *E. coli* and *M. methylotrophus* were grown with shaking at 37 °C overnight. *E. coli* cultures were grown in Luria broth containing 10 μg ml^{-1} tetracycline and *M. methylotrophus* in methanol minimal medium[12] containing 1 μg ml^{-1} tetracycline. Cells were collected by centrifugation and washed in 5 ml of phosphate-buffered saline (PBS). The cells were resuspended in 5 ml of ice-cold PBS and disrupted by sonication. Cell debris was removed by centrifugation at 3,000*g* for 10 min and the cell-free supernatant was used for the assay. Antibodies against ovalbumin were raised in rabbits by injecting into multiple sites 1 mg of ovalbumin (99%; Sigma) emulsified in Freund's complete adjuvant. Injections were repeated once a week for 4 weeks and blood was collected 10–14 days after the last injection. Antibodies were purified by $(NH_4)_2SO_4$ precipitation (15–40% saturation fraction) and dialysis. Polystyrene wells (7 mm diameter × 12 mm deep) were coated with anti-ovalbumin immunoglobin; 300 μl of antibody (10 μg ml^{-1}) in 0.2M $NaHCO_3$ (*p*H 9.2) was pipetted into each well and incubated at 37 °C for 5 h. Wells were washed three times with wash buffer (PBS containing 0.5% normal rabbit serum and 0.1% bovine serum albumin). In triplicate, cell extract (300 μl) was pipetted into wells and the released antigen was allowed to bind overnight at 4 °C. After incubation, wells were washed three times with wash buffer and 300 μl of ^{125}I-anti-ovalbumin immunoglobulin in wash buffer was added (25 ng ml^{-1} ^{125}I-anti-ovalbumin, specific activity 18 μCi μg^{-1}). Wells were incubated at 37 °C for 3 h, washed three times and individual wells counted in a gamma counter. A standard curve was constructed using ovalbumin diluted in PBS or *E. coli* cell extract. As a measure of the immunoreactive ovalbumin-like material produced in *E. coli* and *M. methylotrophus*, the dilution of each strain giving 50% of the maximum binding was taken to be equivalent to the mass of authentic material giving the same counts. The viable cell count per ml of culture and the mass of ovalbumin detected were used to calculate the number of molecules of ovalbumin-like protein produced per cell. ND, not detected.

demonstrated that these strains are also rendered trimethoprim-resistant (our unpublished results).

Chicken ovalbumin, the major egg white protein, is a polypeptide consisting of 386 amino acids. The DNA sequence corresponding to the entire ovalbumin structural gene has been cloned[17,18] and ovalbumin-like protein has been detected in bacterial extracts by immunological methods[19,20].

A *Taq*I fragment of ~2,200 base pairs, containing the entire ovalbumin structural gene, can be excised from the plasmid pOV230 (ref. 19). The *Taq*I ovalbumin fragment was extended with oligo(dC) tails and annealed to pGSS15 DNA to which oligo(dG) sequences had been added after cleavage with *Pst*I. The annealed DNA was used to transform *E. coli* CL499 (*E. coli* CB100 P1 transduced to rk^-mk^+). Ap^{s}Tc^r transformants were screened for the presence of ovalbumin cDNA using a ^{32}P-labelled probe generated by nick-translation[21] of the *Taq*I ovalbumin fragment. Several positive clones from the hybridization were screened for the production of ovalbumin using an immunoradiometric assay. The plasmid, R751 (ref. 22), was then used to mobilize the hybrid plasmids from ovalbumin-producing *E. coli* strains into *M. methylotrophus*. Table 2 shows the levels of ovalbumin expressed in *E. coli* and *M. methylotrophus*. Restriction enzyme mapping of the inserts of clones which did or did not express ovalbumin confirmed that expression was in one orientation only and was probably initiated at the β-lactamase promoter.

In the plasmid pOMP2 (ref. 20), ovalbumin sequences have been fused to part of the *E. coli* β-galactosidase gene. Expression of this hybrid plasmid yields a product in which the first four amino acids of ovalbumin are replaced by the first eight of β-galactosidase and the system is under the control of the strong *lac* UV5 (ref. 23) promoter. This system leads to high levels of expression in *E. coli* and it was of interest to determine whether similar high levels of expression could be achieved in *M. methylotrophus*. The fused β-galactosidase–ovalbumin sequence was isolated from pOMP2 as a single *Hha*I fragment of 2.7 kilobases. pGSS15 was cleaved with *Eco*RI and *Pst*I and the large fragment isolated. The *Hha*I *lac*–ovalbumin and *Eco*RI–*Pst*I pGSS15 fragments were then joined by oligo (dG.dC) tailing. (The construction of these hybrid plasmids is outlined in Fig. 1.) The annealed plasmids were used to transform *E. coli* CL499 and transformants were isolated and analysed in a similar manner to that used for the *Taq*I ovalbumin fragment. Table 2 shows the levels of expression of the *lac*–ovalbumin hybrid sequence in both *E. coli* and *M. methylotrophus*. Restriction enzyme analysis of the plasmid DNA from the hybrids expressing ovalbumin-like protein show that the inserted sequence is expressed in both orientations. However, one orientation (shown in Fig. 1*a*) is favoured (evident in 9 out of 10 isolates); the significance of this unknown. Expression in both orientations, together with the comparable levels of expression of the *lac*–ovalbumin fragment in pGSS15 and in the parental plasmid is evidence that the expression remains under control of the *lac* UV5 promoter.

The levels of immunoreactive ovalbumin detected in *E. coli* from both the β-lactamase and *lac* promoters are ~50-fold lower than expected. The expression of a penicillinase–proinsulin fused protein produced in similar conditions[24] was ~100 molecules per cell and expression of ovalbumin-like protein under *lac* control has been reported at 30,000 molecules per cell[19,20], compared with 2.8 and 760 from the present study. The most likely explanation of these differences concerns the assay systems and in particular the specificity, affinity and avidity of the antibodies used. We have tested this hypothesis using a published radioimmunoassay[20]. The ovalbumin levels detected were similar to those of the immunoradiometric assay described above, indicating that the differences from other published results are probably due to the antisera used.

Although less than in *E. coli*, the level of expression of the two eukaryotic cDNAs in *M. methylotrophus* is significant, indicating that reasonable levels of expression may be achieved in the latter using promoter sequences previously investigated only in *E. coli*. It remains to be determined how closely related these *E. coli* promoter sequences are to the natural *M. methylotrophus* control sequences.

We thank Dr S. N. Cohen for providing pDHFR7, Dr B. W. O'Malley for pOV230, Dr O. Mercereau-Puijalon for pOMP2, Dr B. J. Bachman for CB100 and E. M. Pioli for CL499. We thank Drs M. S. Rose and E. C. Dart and Professor W. J. Brammar for their interest, support and helpful discussions.

Received 30 December 1981; accepted 18 March 1982.

1. Kupper, H. *et al. Nature* **289**, 555–559 (1981).
2. Yelverton, E. *et al. Nucleic Acids Res.* **9**, 731–741 (1981).
3. Keshet, E. *et al. Nucleic Acids Res.* **9**, 19–30 (1981).
4. Goeddel, D. V. *et al. Nature* **287**, 411–416 (1980).
5. Hardy, K., Stahl, S. & Kupper, M. *Nature* **293**, 431–483 (1981).
6. Hitzman, R. A. *et al. Nature* **293**, 719–722 (1981).
7. Meyer, R. J., Figurski, D. & Helinski, D. R. in *DNA Insertion Elements Plasmids and Episomes* (eds Bukhari, A. I., Shapiro, J. A. & Adhya, S. L.) 559–566 (Cold Spring Harbor Laboratory, New York, 1977).
8. Grinter, N. J. & Barth, P. T. *J. Bact.* **128**, 394–400 (1976).
9. Barth, P. T., Tobin, L. & Sharpe, G. S. in *Molecular Biology, Pathogenicity and Ecology of Bacterial Plasmids* (eds Levy, S. B., Clowes, R. C. & Koenig, E. L.) 439–448 (Plenum, New York, 1981).
10. Itakura, K. *et al. Science* **198**, 1056–1063 (1977).
11. Pasek, M. *et al. Nature* **282**, 575–579 (1979).
12. Windass, J. D. *et al. Nature* **287**, 396–401 (1980).
13. Chang, A. C. Y. *et al. Nature* **275**, 617–624 (1978).
14. Meynell, E. & Datta, N. *Nature* **214**, 885–887 (1967).
15. Burchall, J. J. *Ann. N. Y. Acad. Sci.* **186**, 143–152 (1971).
16. Baccanari, D. *et al. Biochemistry* **14**, 5267–5273 (1975).
17. McReynolds, L. A. *et al. Nature* **273**, 723–728 (1978).
18. McReynolds, L. A., Caterall, J. F. & O'Malley, B. W. *Gene* **2**, 217–231 (1977).
19. Fraser, T. H. & Bruce, B. J. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5936–5940 (1978).
20. Mercereau-Puijalon, O. *et al. Nature* **275**, 505–510 (1978).
21. Rigby, P. J. W. *et al. J. molec. Biol.* **113**, 237–251 (1977).
22. Bennett, P. M. & Richmond, M. H. *J. Bact.* **126**, 1–6 (1976).
23. Backman, K. & Ptashne, M. *Cell* **13**, 65–71 (1978).
24. Villa-Komaroff, L. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 3727–3731 (1978).
25. Tabati, H. F. & Flavell, R. A. *Nucleic Acids Res.* **5**, 2321–2332 (1978).
26. Roychoudhury, R., Jay, E. & Wu, R. *Nucleic Acids. Res.* **3**, 101–116 (1976).
27. Lowry, N. J., Rosebrough, N. J., Farr, A. L. & Randall, R. J. *J. biol. Chem.* **193**, 265–270 (1951).

Evidence that a human β-tubulin pseudogene is derived from its corresponding mRNA

C. Deborah Wilde, Carol E. Crowther, T. P. Cripe, Mary Gwo-Shu Lee & N. J. Cowan

Department of Biochemical Sciences, Princeton University, Princeton, New Jersey 08544, USA

Pseudogenes—sequences which are homologous to functional genes but which contain mutational changes precluding the formation of a functional product—seem to be a common feature of eukaryotic genomes. Such sequences were first described in the 5S gene of *Xenopus laevis*[1], and have since been found in several gene families including those of the α- and β-globins of several species[2–10], immunoglobulin V_κ genes[11], the actin genes of *Dictyostelium*[12] and the small nuclear RNA genes of man[13]. Among the globin pseudogenes, there appear to be two kinds of structural organization: genes that have retained their intervening sequences[4–7] and those that have lost them completely[2,3]. A possible explanation for the generation of such intron-lacking genes involves the reverse transcription of a processed mRNA to form cDNA, and the introduction of this cDNA copy into the genome either by recombinant heteroduplex formation[2], insertion into a staggered chromosomal break[3], or via a retrovirus intermediate[14]. Here we report the complete sequence of a human β-tubulin pseudogene. The sequence data reveal the absence of any intervening sequences, and the presence, in the genome, of an uninterrupted 17 base-pair (bp) tract of A residues 14 base-pairs 3′ to a poly(A) signal (AATAAA)[15]. This sequence organization corresponds closely to the sequence organization of the poly(A) signal and poly(A) tract in a β-tubulin mRNA[16].

Screening of two human recombinant bacteriophage libraries with chicken α- and β-tubulin cDNA probes has resulted in the isolation of most of the tubulin-like sequences contained in the human genome[17–19]. Structural analysis of the β-tubulin sequences has revealed genes of varying length, including some that are close to, or apparently less than the size required to encode a functional mRNA[17]. The complete sequence of one of these genes, determined by the stratagem shown in Fig. 1, is given in Fig. 2. There is a striking overall homology between the chicken and human sequences within most of the coding sequence: 90% of the amino acid sequence is unchanged, while 82% of the bases are conserved. Among those bases that differ, 75% are 'silent' substitutions that do not affect the amino acid sequence. Most of these substitutions are from G or C to A or T residues, accounting for the lower (56%) G+C content of the human relative to the chicken (62%) gene within the coding sequence. The human gene (46β) contains no intervening sequences. It has a 3′ untranslated region which is 44 nucleotides shorter than the corresponding chicken sequence, and is largely non-homologous. Immediately upstream from the initiator ATG in 46β, the sequence is markedly A+T rich (63%), in contrast to the chicken 5′ untranslated region (21%).

Two features render the human sequence incapable of yielding a functional mRNA. First, at amino acid position 230, the serine codon TCG appears as a termination codon (TAA); second, at amino acid position 270, the third base of the phenylalanine codon (TTC) is absent, resulting in a frameshift leading to a termination codon (TGA) beginning at the second base of amino acid 299. In addition, an unexpected feature of the human sequence is the presence, 14 bp 3′ to the poly(A) signal (AATAAA)[15], of an uninterrupted tract of 17 A residues. The spacing between the poly(A) signal and the oligo(A) tract corresponds very closely to the spacing between the chicken cDNA poly(A) signal and its poly(A) sequence. Moreover, the human and chicken sequences share a common trinucleotide (CTT) and pentanucleotide (TTTGT) in this region. This sequence organization strongly suggests that the 46β human β-tubulin pseudogene arose by reverse transcription from a processed human β-tubulin gene transcript.

A second remarkable feature of the 46β human pseudogene is the existence of a 11-bp sequence 3′ to the poly(A) tract that is directly repeated 170 bp 5′ to the initiator ATG (Fig. 2). Because the sequence of a human β-tubulin mRNA has not been determined, it is not possible to identify the equivalent of the cap site in 46β. However, the location of the repeat sequences is consistent with their generation by repair of a staggered chromosomal break following insertion of the cDNA into the chromosome.

Sequence analysis of a second human β-tubulin gene reveals the occurrence of an in-phase termination codon at amino acid position 361, and the apparent absence of sequences encompassing the entire 5′ end up to amino acid position 54 (C.E.C., C.D.W., T.P.C., M.G.-S.L. and N.J.C., unpublished results). In contrast to the data reported here, however, this pseudogene contains a short intervening sequence with correct consensus splice signals. Thus, in common with the globin pseudogenes[2–10], human β-tubulin pseudogenes have arisen with and without intervening sequences. This observation, and the existence of pseudogenes with or without abutting direct repeat sequences[13], suggests that they were produced by various independent and probably unrelated mechanisms. Gene duplication, with a subsequent loss of selective pressure on one resulting copy, can satisfactorily explain the origin of intron-containing pseudogenes. These sequences could remain linked to their functional counterpart[4,7–9] or be transferred to a new location by chromosomal translocation, copy insertion or as part of a transposable element. On the other hand, the integration of cDNA copies might occur either directly, or via a retrovirus intermediate, in either case resulting in the random dispersion of intron-lacking pseudogenes throughout the genome.

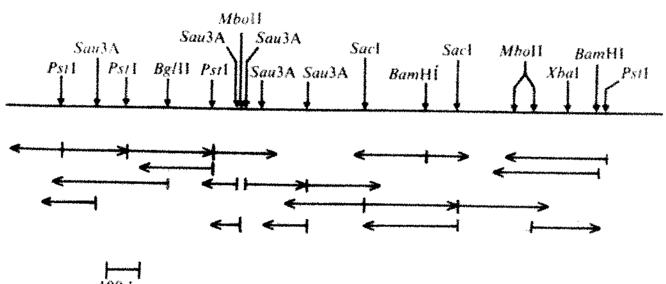


Fig. 1 Sequencing strategem for clone 46β. The sequencing method used was that of Sanger *et al.*[20]. The large *Pst*I fragment was subcloned by digestion of the entire 46β λ clone[17] with *Pst*I, and ligation of the resulting fragments into pBR322. The appropriate subclone was identified by the method of Grunstein and Hogness[21] using a chicken β-tubulin cDNA probe[22] labelled with ^{32}P by nick-translation[23]. This subclone, or the entire 46β λ clone, was used as a substrate for digestion with the restriction enzymes shown. In each case, the resulting fragments were subcloned into the replicative form of M13 mp 8 and/or mp 9. Plaques containing DNA to be sequenced were identified on nitrocellulose replicas[24] by hybridization with the chicken β-tubulin cDNA probe. Sequence analysis revealed two small *Pst*I fragments containing the 5′ end of the 46β-tubulin gene. These fragments were not detected in previous mapping experiments[17] because of their small size and the limitations of blot transfer procedures.

0028–0836/82/180083—02$01.00

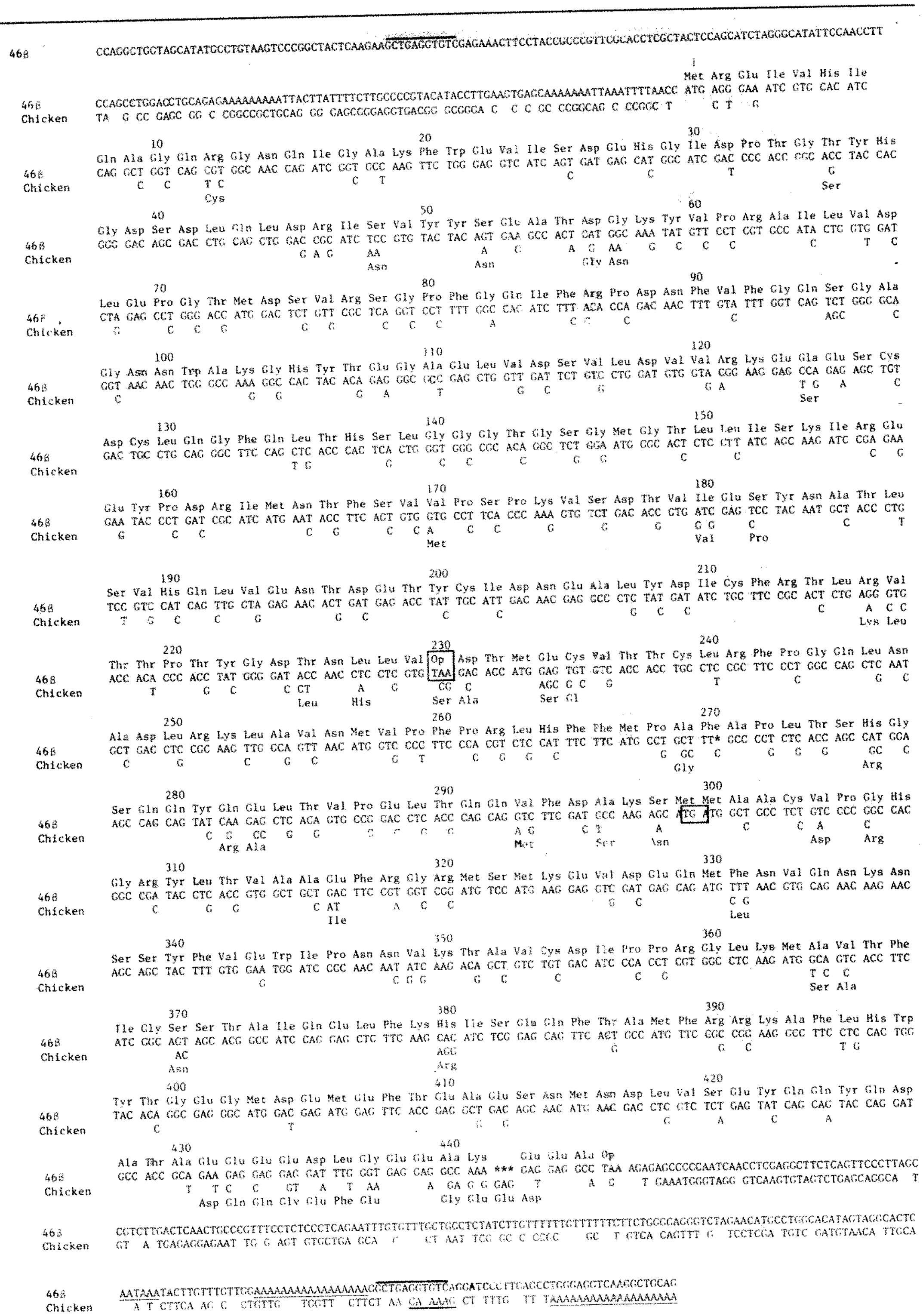


Fig. 2 Sequence of the human β-tubulin pseudogene contained in clone 46β. The chicken β-tubulin cDNA sequence[16] is shown for comparison. Homologies with the chicken cDNA sequence were identified using the computer program of Staden[25]. A termination codon appears at amino acid position 230 (boxed). In addition, a frameshift resulting from a single base deletion at amino acid position 270 has produced an in-phase termination codon (boxed). The poly(A) signals[15] and 3' oligo(A) tracts are underlined. A flanking 11 base-pair direct repeat is overlined. *, Deleted base.

This work was supported by NIH grant GM 26456. C.D.W. acknowledges the award of a Muscular Dystrophy Association postdoctoral fellowship.

Received 27 January; accepted 24 February 1982.

1. Jacq, D., Miller, J. R. & Brownlee, G. G. *Cell* **12**, 109–120 (1977).
2. Nishioka, Y., Leder, A. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2806–2809 (1980).
3. Vanin, E. F., Goldberg, G. I., Tucker, P. W. & Smithies, O. *Nature* **286**, 222–226 (1980).
4. Jahn, C. L. *et al. Cell* **21**, 159–168 (1980).
5. Lacy, E. & Maniatis, T. *Cell* **21**, 545–553 (1980).
6. Proudfoot, N. J. & Maniatis, T. *Cell* **21**, 537–554 (1980).
7. Cleary, M. L., Schon, E. A. & Lingrel, J. B. *Cell* **26**, 181–190 (1981).
8. Lacy, E., Hardison, R. C., Quon, D. & Maniatis, T. *Cell* **18**, 1273–1283 (1979).
9. Proudfoot, N. J., Shander, M. H., Manley, J. L., Gefter, M. & Maniatis, T. *Science* **209**, 1329–1336 (1980).
10. Leder, A., Swan, D., Ruddle, F., D'Eustachio, P. & Leder, P. *Nature* **293**, 196–200 (1981).
11. Bentley, D. L. & Rabbitts, T. H. *Nature* **288**, 730–733 (1980).
12. Firtel, R. A., Timon, R., Kimmel, A. R. & McKeown, M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 6206–6210 (1979).
13. Van Arsdell, S. W. *et al. Cell* **26**, 11–17 (1981).
14. Goff, S. P., Gilboa, E., Witte, O. N. & Baltimore, D. *Cell* **22**, 777–785 (1980).
15. Proudfoot, N. J. & Brownlee, G. G. *Nature* **263**, 211–214 (1976).
16. Valenzuela, P. *et al. Nature* **289**, 650–655 (1981).
17. Cowan, N. J., Wilde, C. D., Chow, L. T. & Wefald, F. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4877–4881 (1981).
18. Wilde, C. D., Chow, L. T., Wefald, F. & Cowan, N. J. *Proc. natn. Acad. Sci. U.S.A.* **79**, 96–100 (1982).
19. Wilde, C. D., Crowther, C. E. & Cowan, N. J. *J. molec. Biol.* (in the press).
20. Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. *J. molec. Biol.* **143**, 161–178 (1980).
21. Grunstein, M. & Hogness, D. S. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3961–3965 (1975).
22. Cleveland, D. W. *et al. Cell* **20**, 95–195 (1980).
23. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. *J. molec. Biol.* **113**, 237–251 (1977).
24. Benton, W. D. & Davis, R. W. *Science* **196**, 180–181 (1977).
25. Staden, R. *Nucleic Acids Res.* **8**, 3673–3694 (1980).

BOOK REVIEWS

The illusory quest for primary colours

J. D. Mollon

Colour Vision in the Nineteenth Century: The Young–Helmholtz–Maxwell Theory. By Paul D. Sherman. Pp.233. ISBN 0-85274-376-9. (Adam Hilger/Heyden: 1981.) £35, $77.

THERE is nothing like colour vision for sorting out the men from the boys. Over the past three centuries the field has attracted the strongest minds — Newton, Thomas Young, Helmholtz, Maxwell — and it is an instructive exercise for the modern student to recapitulate the successive insights of these giants. Equally, however, colour vision has always attracted more than its share of heretics and amateurs, perhaps because colour is so immediate a part of everyone's private experience. In his new book P.D. Sherman traces the sequence of development of the trichromatic theory — and describes too the diversionary actions fought by heretics along the way. Finally, however, his own conceptual development does not keep up and, somewhere around the middle of the nineteenth century, he is himself left behind.

The sub-title is a better guide to the book's contents than is the title: the text is almost exclusively confined to trichromatic aspects of colour vision and it covers the period 1700–1860. There is a quite detailed account of the physical trichromatists of the eighteenth century, whereas there is nothing about the developments of the last four decades of the nineteenth century, such as Helmholtz's line-element, Hering's opponent-process theory and König's derivation of fundamental sensitivities that are very close to those of the photoreceptors.

After discussing the historical position of Thomas Young, Dr Sherman devotes a chapter to one of the most influential heretics in the history of colours, David Brewster, who maintained as late as 1850 that there were only three physical forms of light. Dr Sherman describes the experiments that led Brewster to this position and explains how they were finally answered by Helmholtz. He gives a further chapter to one of the neglected heroes of colour science, Grassmann, who introduced the vector representation of colours. Other chapters are devoted to colour blindness and to geometrical representations of colour, the latter illustrated by several colour plates. This period of colour science is covered in greater detail than in any existing text and useful reference lists are provided.

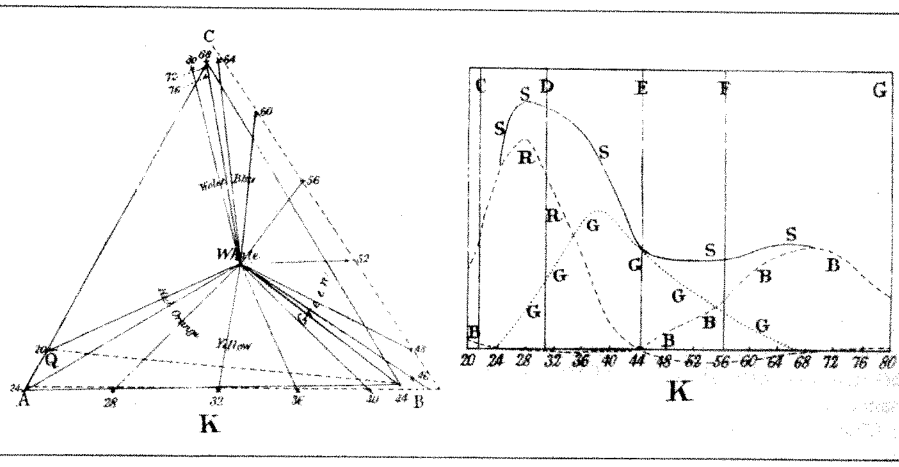
The final two chapters are essentially précis of the great papers on colour vision written by Clerk Maxwell in the period 1855–1860. On the whole the account is admirably accurate and clear (though the student would do as well to read the masterly originals), but when Dr Sherman strays from the path of strict summary he stumbles, and at two crucial points shows that he does not really understand colour vision. The first difficulty centres on the concept of an imaginary primary, a light that has the property of stimulating only a single class of photoreceptors. Since the three classes of receptor have broad and overlapping spectral sensitivities, there do not exist real lights with this property, yet Dr Sherman sometimes suggests that the long-wave fundamental corresponds to an actual wavelength, and he regrets aloud that Maxwell did not measure it. More seriously, he goes on to confuse Maxwell's spectral mixture curves with the actual spectral sensitivities of the receptors. The curves shown by Maxwell represent the proportion of each of his three "standard" wavelengths required to match any other spectral light, but Dr Sherman writes: "Each curve, being an individual physiological response, shows how each primary stimulates a perceived sensation at each point of the spectrum". In fact, Maxwell's curves are very different from the spectral sensitivities of the photoreceptors; but here Dr Sherman leaves us just as the problem is getting interesting.

But was Maxwell himself clear on these two related issues? Certainly on the first question he seems twice to have changed his mind. In his first paper on colour (1855) he writes, unambiguously:

> Though the homogeneous rays of the prismatic spectrum are absolutely pure in themselves, yet they do not give rise to the "pure sensations" of which we are speaking. Every ray of the spectrum gives rise to all three sensations, though in different proportions . . . .

And in his last paper (1871) he writes: ". . . the true primary red is not exactly represented in colour by any part of the spectrum". In 1860, however, Maxwell tabulated the wavelengths of three "primary colours", adding "I have strong reason to believe that these are the three primary colours corresponding to three modes of sensation in the organ of vision . . ." (*Phil. Trans. R. Soc.*, Vol.150). On the interpretation of spectral mixture curves, his words are more guarded than those of Dr Sherman.

So was Maxwell confused? This is a nice historical question, but I think the answer is probably no, for the following reasons. First, he was working with a version of the trichromatic theory which differs from the modern one: he postulated three primary sensations each corresponding to the activity of an independent class of receptor. We now know that the psychophysical transformation is more



Colour mixing results obtained by Clerk Maxwell for his wife, Katherine. In both diagrams, the numbers 20 to 80 represent arbitrary scale readings on Maxwell's apparatus but correspond to various wavelengths, running from red to blue. By applying a centre-of-gravity rule to the chromaticity diagram on the left, one can find what monochromatic light or spectral mixture will match the colour of a given mixture of three standard wavelengths (say, those wavelengths numbered 24, 44, 68). The 'spectral mixture curves' (R,G,B) on the right show directly the intensities of the three standard colours required to match any given wavelength in the spectrum.

complicated: for example, the "red-sensitive" cones are maximally sensitive in the yellow-green and what is maximal at longer wavelengths is the *ratio* of absorptions in the long- and middle-wavelength cones. Second, given his theory and his actual results, Maxwell had empirical reasons for supposing that there existed three real wavelengths each of which produced a nearly pure sensation; for his results, when plotted in his chromaticity space, suggested that the spectrum locus consisted simply of two straight lines that intersected in the green and thus he was led by his data to suppose that a light placed at any apex of the resulting triangle would come close to stimulating only one sensation.

Maxwell's treatment of colour deserves a more careful assessment than is possible here; but much of the problem has escaped Dr Sherman. The one single error that most held back the development of the trichromatic theory was the belief — still to be heard today — that the peak sensitivities of the photoreceptors must correspond to "primaries", either phenomenologically pure "primary hues" or those primaries in a colour-matching experiment that, by direct mixing, allow the matching of the largest gamut of other colours. Only after Maxwell was it realized that the latter primaries should be lights that produce the highest ratio of absorptions in different classes of receptor. Dr Sherman's useful book would have been so much the more valuable if it had identified the essential ambiguity of the term "primary colour" in the literature of the mid-nineteenth century. More generally, his readers would have gained if he had much more clearly set out, in an early chapter, the modern distinction between a chromaticity space (representing physical mixtures that produce matches) and spaces in which colours are so placed as to be uniformly spaced phenomenologically. The mapping of one to the other is now known to be very complex.

The historian of science has a peculiarly difficult trade: he must be master of his period and of at least four languages, and he must be also master of the science. If he is to show us the limits of his subjects' thoughts, he must have a conceptual advantage over them and must share this advantage with his reader. He must be Whig without being presentist: the two are not the same. □

John Mollon is a Lecturer in Experimental Psychology at the University of Cambridge.

François Jacob on possible worlds

P.B. Medawar

The Possible and the Actual. By François Jacob. Pp.70. Hbk ISBN 0-295-95888-X; pbk ISBN 0-394-70671-4. (University of Washington Press/Pantheon: 1982.) Hbk \$8.95; pbk \$3.95.

THE antithesis embodied in the title of Jacob's short, brilliant and beautifully written little essay can be taken in two ways. In *The Nature of the Physical World*, Sir Arthur Eddington wrote:

> I need scarcely add that the contemplation in natural science of a wider domain than the actual leads to a far better understanding of the actual.

An important fraction of Jacob's essay is indeed an exercise of just this kind. It is an exercise more easily possible in the biological than in the physical sciences because in the latter, as Jacob remarks in a Kantian aside, the right-seemingness and naturalness of the natural world may to some extent reflect "the very way our brains work". In biology, however, it is not a tremendous feat to imagine things otherwise than they really are. Why sex, for example? — and why two sexes and not three? The causal-explanatory answer is not easy but the teleological answer that sexuality is "a diversity-generating device" is satisfying and makes sense of what would otherwise be a puzzle.

The second sense in which we may read Jacob's antithesis is that in which all scientific enquiry is a duologue between the possible and the actual, between what *might* be true and what is in fact the case.

This conception, in which I myself am a firm believer, is sufficiently permissive to make room for myth as an important element in man's attempt to explain his condition, for "whether mythic or scientific, the view of the world that man builds is always largely a product of his imagination". Reflection on this theme allows Jacob to divagate interestingly and amusingly on myths that bear upon sexuality. But of course myths, however effectively they may appease curiosity or dispel the disquiet of ignorance, always fail that difficult exam in which our imaginings are measured up against reality. Yet even in the most "scientific" explanations today there is just a faint hint of the explanatory glibness of the myth — and it is for this reason, I think, that Jacob is a little reserved about natural selection: he does not believe in the universal explanatory value of any one formula.

Needless to say, no one with any understanding of the intellectual tradition to which Jacob belongs and of which he is so able a spokesman could entertain even momentarily the suspicion that Jacob himself is in any serious sense a mystic.

It is an added merit of this book that there is no Message, and that it is altogether too polished and urbane to harangue us at any time. It is a very scientific book which embodies more neatly than I have seen expressed elsewhere the conviction of most scientists that

> . . . there are many possible worlds; but the interesting one is the world that exists and has already shown itself to be at work for a long time. Science attempts to confront the possible with the actual. It is the means devised to build a representation of the world that comes ever closer to what we call reality. □

Sir Peter Medawar is Head of the Transplantation Biology Section at the Medical Research Council's Clinical Research Centre, Harrow, Middlesex.

Marine microfossils

B. M. Funnell

Microfossils from Recent and Fossil Shelf Seas. Edited by J. W. Neale and M. D. Brasier. Pp.380 UK ISBN 0-85312-338-1; US ISBN 0-470-27220-1. (Ellis Horwood/ Halsted:1981.) £35, \$76.65.

A PRINCIPAL merit of this book is the variety of topics it contains. In 28 chapters it covers every geological period from the Silurian to the Recent; a geographical range embracing Spitsbergen and the Barents Sea, East Africa, the equatorial Pacific, the western North Atlantic, Israel, Iran, Pakistan, North-West Europe and the British Isles; and microfossil groups as varied as conodonts, algae, acritarchs, dinoflagellates, ostracods, benthonic and planktonic foraminifers and radiolaria. The organizers of the symposium at Hull University in July 1980, from which this volume derives, were indeed fortunate in having such a wide diversity of contributions on which to draw.

Another successful feature of the book is the high degree of uniformity of contributions, none of which are either too long or too ambitious, and only one of which is reduced to extended abstract dimensions.

The volume as a whole is arranged in stratigraphical order. Approximately 15% is devoted to the Palaeozoic, 35% to the Mesozoic, 20% to the Tertiary, and 30% to the Quaternary and Recent. Most articles apply themselves to stratigraphical or palaeoenvironmental interpretation. No fundamentally new approaches are introduced, and the methodologies used are essentially those developed in the 1960s and early 1970s. A particularly disappointing aspect is the poor understanding of physical and chemical oceanography exhibited by some authors. A few contributions begin to break through this deficiency, for example that dealing with anoxic events in the late Cretaceous and the salutary study of microfossil transport in a modern estuary;

but others, on equatorial Pacific palaeotemperatures and possibly Palaeozoic shelf bathymetries, for instance, tend simply to produce an "acceptable" interpretation that is derivative, and can scarcely be sustained on the microfossil evidence presented.

In spite of these shortcomings the book provides a good review of current research in marine micropalaeontology as practised by members of the British Micropalaeontological Society and their international associates. The inclusion of chapters on the Permian of Iran, the Eocene of Israel and the Recent of East Africa goes a long way towards redressing what would otherwise be a strong emphasis on temperate zone, North Atlantic and North European occurrences.

This collection of papers should serve a useful purpose in stimulating renewed interest in this currently most useful, but sometimes neglected branch of the earth and biological sciences. It can certainly be recommended for reading by both advanced undergraduates and specialist research workers in the field. □

B. M. Funnell is Professor of Environmental Sciences at the University of East Anglia.

Muscle laid bare to the bone

R.M. Simmons

The Structural Basis of Muscular Contraction. By John Squire. Pp.716. ISBN 0-306-40582-2. (Plenum: 1981.) £34.65, $55.

THE publication of an advanced, or even intermediate, text on muscle is a rare event, an obvious disincentive being that the field is now too big and in many areas too technical to be encompassed satisfactorily in a single book or by a single author. The most one can hope for is an expert treatment of a particular area, and that is what John Squire has done for muscle structure.

Progress in understanding the details of the structure of muscle has depended heavily on the use of X-ray diffraction and on optical diffraction from electron micrographs. The author recognizes the difficulty that diffraction theory presents to non-experts and includes in the introductory material a lucid non-mathematical treatment together with examples of common structures found in proteins and polypeptides. There is also a useful description of the methods used in electron microscopy. Thereafter it is a straight slog through the static structure of muscle, the various isolated and assembled preparations from individual molecules through paracrystals, thick and thin filaments, to superlattices, all 300 pages of it. Such an account could easily have ended up as no more than a useful compendium of techniques and results for the expert, but Squire succeeds on the whole in pursuing what is relevant to understanding structure and function in real muscles. Some readers may find that the author's enthusiasm for research on thick filament assembly exceeds their own, and that in the same area he is a little over-protective towards his own models; there is more than a hint of special pleading for smooth muscle myosin ribbons and the boot goes in a bit heavily into the opposition on vertebrate muscle filaments.

The general outlines of most of the static structures emerge with some clarity. But no question is settled until it is settled right, and that does not come easily in muscle research. In spite of years of patient grind we still do not know the precise arrangement of the cross-bridges on the outside of the thick filament in relaxed muscle, let alone the packing of the tail portions in the backbone. Doubts have been raised about the steric blocking mechanism of tropomyosin on the thin filament in the regulation of contraction. There are still arguments about the interpretation of the clearest static picture of the intact lattice — insect fibrillar muscle in rigor. The author gives grounds for believing that many of the static structural problems will be solved and indeed there is some confirmation for this view from results too recent to have been included in the book, such as new reconstructions of decorated actin filaments and improved staining methods for cross-bridges in isolated thick filaments.

When it comes to solving the main problem in muscular contraction, of what are the dynamic structural changes underlying force generation and work production, and of how these are related to biochemical events, the grounds for optimism are fewer (and indeed there are fewer pages on it in the book). The author's profession of faith as a molecular biophysicist is that the more one knows about a structure, the more one grows capable of saying how it works. But it is not at all clear that a detailed knowledge of static structure of the filaments or even of the arrangement of cross-bridges in the relaxed and rigor states is of much help in understanding dynamic changes. Contracting muscle displays a high degree of disorder and although improvements in techniques such as time-resolved X-ray diffraction have led to some suggestions about cross-bridge action, there is not much structural information to go on. One has to conclude that a satisfactory solution lies some way ahead and that new techniques will be needed to find it. □

R.M. Simmons is in the MRC Cell Biophysics Unit, King's College, University of London.

BOOKS RECEIVED

Astronomy

HUNT, G. (ed.) Uranus and the Outer Planets. Proceedings of the IAU/RAS Colloquium, No. 60. Pp.307. ISBN 0-521-24573-7. (Cambridge University Press: 1982.) £18.

JASCHECK, M. and GROTH, H-G. (eds). ß-Stars. Symposium 98, Munich, April 6–10. Pp.523. Hbk ISBN 90-277-1366-9; pbk ISBN 90-277-1367-7. (Reidel: 1982.) Hbk Dfl.130.00, $56.50; pbk Dfl.65, $28.50.

PRESIDENCIA DEL GOBIERNO. DIRECCION GENERAL DEL INSTITUTO GEOGRAFICO NACIONAL. Anuario del Observatorio Astronomico de Madrid Para 1982. Pp.449. Pbk ISBN 84-500-3470-1958. (Talleres Del Instituto Geografico: 1981.) Pbk np.

Physics

ALFERNESS, R. C. *et al.* 1st European Conference on Integrated Optics, September 1981. IEE Conference Publication, No. 201. Pp.112. (no ISBN) (Institution of Electrical Engineers, Stevenage, Herts: 1981.) Pbk £19.25 (UK) $55 (Americas) £22 (elsewhere).

ALLEN, P. D. *et al.* International Conference on Advanced Infrared Detectors and Systems, October 1981. IEE Conference Publication, No. 204. Pp.138. Pbk ISBN 0-85296-9. (Institution of Electrical Engineers, Stevenage, Herts: 1981.) £18 (UK) $52.50 (Americas) £21 (elsewhere).

BARRETT, H. H. and SWINDELL, W. Radiological Imaging. The Theory of Image Formation, Detection and Processing, Vol. 2. Pp.693. ISBN 0-12-079602-3. (Academic: 1981.) $55.

BOOKER, H. G. Energy in Electromagnetism. Pp.360. ISBN 0-906048-59-1. (Institution of Electrical Engineers, London & New York: 1982.) £25 (UK) $75 (Americas) £30 (elsewhere).

CASSELS, J. M. Basic Quantum Mechanics. 2nd Edn. Pp.205. Hbk ISBN 0-333-18599-4; pbk ISBN 0-333-31768-8. (Macmillan Press, London & New York: 1982.) Hbk £22; pbk £9.95.

GEKKER, I. R. The Interaction of Strong Electromagnetic Fields with Plasmas. Pp.324. ISBN 0-19-851467-0. (Oxford University Press: 1982.) £35.

GREEN, M. A. Solar Cells: Operating Principles, Technology, and System Applications. Pp.274. ISBN 0-13-822270-3. (Prentice-Hall: 1982.) £20.95.

HUDSON, J. E. Adaptive Array Principles. IEE Electromagnetic Waves Series, 11. Pp.253. ISBN 0-906048-55-9. (Institution of Electrical Engineers, London & New York: 1981.) £26.50 (UK) $77 (Americas) £31 (elsewhere).

ISHAM, C. J., PENROSE, R. and SCIAMA, D. W. (eds). Quantum Gravity, 2. A 2nd Oxford Symposium. Pp.669. ISBN 0-19-851952-4. (Oxford University Press: 1981.) £28.

JAMES, J. R., HALL, P. S. and WOOD, C. Microstrip Antenna. Theory and Design. IEE Electromagnetic Waves Series, 12. Pp.290. ISBN 0-906048-57-5. (Institution of Electrical Engineers, London & New York: 1981.) £26.50 (UK) $77 (Americas) £31 (elsewhere).

JULL, E. V. Aperture Antennas and Diffraction Theory. IEE Electromagnetic Waves Series, 10. Pp.173. ISBN 0-906048-52-4. (Institution of Electrical Engineers, London & New York: 1981.) £23.75 (UK) $68 (Americas) £27 (elsewhere).

KRAUSE, E. (ed.). Advances in Fluid Mechanics. Lecture Notes in Physics, Vol. 148. Proceedings of a Conference held at Aachen, March 1980. Pp.361. ISBN 3-540-11162-X. (Springer-Verlag: 1981.) DM 43, $20.10.

MICHEL, R., COUSTEIX, J. and HOUDEVILLE, R. (eds). International Union of Theoretical and Applied Mechanics. Unsteady Turbulent Shear Flows. Symposium, Toulouse, France, May 1981. Pp.424. ISBN 3-540-11099-2. (Springer-Verlag: 1981.) DM89, $41.50.

MOSS, T. S. and STENHOLM, S. (eds). Progress in Quantum Electronics, Vol. 6. Pp.305. ISBN 0-08-028387-X. (Pergamon: 1981.) £31.50, $72.50.

OBERHOFER, M. and SCHARMANN, A. (eds). Ispra Courses. Applied Thermoluminescence Dosimetry. Lectures of a Course held at the Joint Research Centre, Ispra, Italy, November 1979. Pp.396. ISBN 0-84274-544-3. (Adam Hilger, Bristol: 1981.) £35.

PRESS, W. Single-Particle Rotations in Molecular Crystals. Springer Tracts in Modern Physics, Vol.92. Pp.129. ISBN 3-540-10897-1. (Springer-Verlag: 1981.) DM 53, $24.70.

RAINE, D.J. and HELLER, M. The Science of Space–Time. Pp.244. ISBN 0-912918-12-8. (Pachart, Arizona: 1981.) Np.

SKWIRZYNSKI, J. K. (ed.). Theoretical Methods for Determining the Interaction of Electromagnetic Waves with Structures. NATO Advanced Study Institutes Series E: Applied Sciences, No. 40. Pp.913. ISBN 90-286-2161-X. (Sijthoff & Noordhoff, The Netherlands: 1981.) Np.

STRAUMANN, N. Allgemeine Relativitästheorie und Relativistische Astrophysik. Lecture Notes in Physics, Vol. 150. Pp.418. Flexi ISBN 3-540-11182-4. (Springer-Verlag: 1981.) Flexi DM41, $19.10.

TELLER, E. (ed.). Magnetic Confinement Part B. Fusion, Vol.1. Pp.529. ISBN 0-12-685241-3. (Academic: 1981.) $50.

ULABY, F. T., MOORE, R. K. and FUNG, A. K. Microwave Remote Sensing Fundamentals and Radiometry. Microwave Remote Sensing Active and Passive, Vol. 1. Pp.456. ISBN 0-201-10759-7. (Addison-Wesley: 1981.) Np.

WAIT, J. R. (Lectures on) Wave Propagation Theory. Pp.348. Hbk ISBN 0-08-026345-3; flexi ISBN 0-08-026344-5. (Pergamon: 1981.) Hbk £20, $40; flexi £9, $20.

WILL, C. M. Theory and Experiment in Gravitational Physics. Pp.342. ISBN 0-521-23237-6 (Cambridge University Press: 1982.) £37.50.

WILLARDSON, R.K. and BEER, A.C. (eds). Semiconductors and Semimetals. Vol.16, Defects, (HgCd)Se, (HgCd)Te. Pp.266. ISBN 0-12-752116-X. (Academic: 1981.) $37.50.

Chemistry

ANDERS, E. *et al.* Topics in Current Chemistry, Vol. 99. Pp.133. ISBN 3-540-10920-X. (Springer-Verlag: 1981.) DM68, $31.70.

BAILEY, P. S. Ozonation in Organic Chemistry. Vol. 1, Olefinic Compounds. Organic Chemistry, Vol. 39. Pp.272. ISBN 0-12-073101-0. (Academic: 1981.) $30.50.

BAMBAKIDIS, G. (ed.). Metal Hydrides. NATO Advanced Study Institutes Series, Vol.76. Proceedings of a NATO Advanced Study Institute held June 1980 in Rhodes, Greece. Pp.385. ISBN 0-306-40891-0. (Plenum: 1981.) $49.50.

CLARKE, F. H. Calculator Programming for Chemistry and the Life Sciences. Pp.226. ISBN 0-12-175320-4. (Academic: 1981.)

CLARKE, M.J. *et al.* (eds). Structure and Bonding, Vol.47. Pp.126. ISBN 3-540-10788-6. (Springer-Verlag: 1981.) DM 64, $29.80.

COLVIN, R.W. Silicon in Organic Synthesis. Pp.348. Hbk ISBN 0-408-10619-0; pbk ISBN 0-408-10831-2. (Butterworths: 1981.) Hbk £30; pbk £15.

FERRARO, J. R. and BASILE, L. J. (eds). Fourier Transform Infrared Spectroscopy. Applications to Chemical Systems, Vol. 1. Pp.311. ISBN 0-12-254101-4. (Academic: 1981.) $34.

GAJEWSKI, J.J. Hydrocarbon Thermal Isomerizations. Organic Chemistry, Vol.5. Pp.442. ISBN 0-12-273350-9. (Academic: 1981.) $45.

HAN, C. D. Multiphase Flow in Polymer Processing. Pp.459. ISBN 0-12-322460-8. (Academic: 1981.) $58.50.

HAWKINS, D.T. (ed.). Chemical Vapor Deposition 1969–1980. A Bibliography. Pp.737. ISBN 0-306-65201-3. (IFI/Plenum: 1981.) $115.

HILL, G. C. and HOLMAN, J. S. Chemistry in Context: Laboratory Manual and Study Guide. Pp.240. Flexi ISBN 0-17-448059-8. (Thomas Nelson: 1982.) £6.35.

HUTZINGER, O. (ed.). The Handbook of Environmental Chemistry, Vol. 2, Part B. Reactions and Processes. Pp.205. ISBN 3-540-11107-7/0-387-11107-7. (Springer-Verlag: 1982.) DM98, $45.70.

JOHNSON, D. A. Some Thermodynamic Aspects of Inorganic Chemistry. 2nd Edn. Pp.282. Hbk ISBN 0-521-07108-9; pbk ISBN 0-521-28521-1. (Cambridge University Press: 1982.) Hbk £18; pbk £6.95.

KATRITZKY, A. R. and BOULTON, A. J. (eds). Advances in Heterocyclic Chemistry, Vol. 29. Pp.405. ISBN 0-12-020629-3. (Academic: 1981.) $62.50.

LARSSON, R. (ed.). Perspectives in Catalysis. In Commemoration of Jöns Jacob Berzelius. Proceedings of the 12th Swedish Symposium, October 1979 including the 1st Berzelius Lecture in Catalysis. Pp.122. Pbk ISBN 91-40-04783-0. (CWK Gleerup: 1981.) Pbk np.

LAWRENCE, J.F. Organic Trace Analysis by Liquid Chromatography. Pp.288. ISBN 0-12-439150-8. (Academic: 1981.) $34.

LÖWDIN, P.-O. (ed.). Advances in Quantum Chemistry, Vol. 13. Pp.251. ISBN 0-12-034813-6 (Academic: 1981.) $42.

LÖWDIN, P.-O. (ed.). Advances in Quantum Chemistry, Vol. 14. Pp.248. ISBN 0-12-034814-4. (Academic: 1981.) $46.50.

MARINSKY, J. A. and MARCUS, Y. (eds). Ion Exchange and Solvent Extraction. A Series of Advances, Vol. 8. Pp.464. ISBN 0-8247-1333-8. (Dekker: 1981.) SwFr154.

RYLES, A. P., SMITH, K. and WARD, A. P. Worked Examples in Essential Chemistry. Pp.161. Hbk ISBN 0-471-27972-2; pbk ISBN 0-471-27975-7. This book is a companion text to Essential Organic Chemistry. (Wiley: 1982.) Hbk £11, $26.50; pbk £4.40, $9.95.

STJÄRNE, L *et al.* (eds). Chemical Transmission 75 Years. Based on Proceedings of the 2nd Nobel Conference held at the Wenner-Gren Centre, Stockholm, December 1980. Pp.562. ISBN 0-12-671480-0. (Academic: 1982.) £43. 40, $89.

SVAROVSKY, L. (ed.). Solid–Liquid Separation. 2nd Edn. Pp.556. ISBN 0-408-70943-X. (Butterworths: 1981.) £30.

TEWARI, P.H. (ed.). Adsorption From Aqueous Solutions. Proceedings of a Symposium held March, 1980 as a Satellite Symposium to the Meeting of the American Chemical Society, Houston, Texas. Pp.248. ISBNZ 0-306-40747-7. (Plenum: 1981.) $35.

VON GUSTORF, E.A.K., GREVELS, F.-W. and FISCHLER, I. (eds). The Organic Chemistry of Iron. Vol.2, Organometallic Chemistry. Pp.340. ISBN 0-12-417102-8. (Academic: 1981.) $39.

Technology

ELISEEVA, V.I. *et al.* Emulsion Polymerization and Its Applications in Industry. Pp.225. ISBN 0-306-10961-1. (Consultants Bureau, New York and London: 1981.) $49.50.

GOLDSTEIN, J.I. *et al.* Scanning Electron Microscopy and X-Ray Microanalysis. A Text for Biologists, Materials Scientists, and Geologists. Pp.673. ISBN 0-306-40768-X. (Plenum: 1981.) $29.50.

MORGAN, R.P. and ICERMAN, L.J. *et al.* Renewable Resource Utilization for Development. Pp.394. ISBN 0-08-026338-0. (Pergamon: 1981.) £19.75, $39.50.

OLIVER, J.F. (ed.). Adhesion in Cellulosic and Wood-Based Composites. NATO Conference Series VI: Materials Science, Vol.4. Proceedings of a Conference held May 1980 at Queen's University, Kingston, Ontario, Canada. Pp.261. ISBN 0-306-40812-0. (Plenum: 1981.) $35.

PIERCE, J.R. Signals: The Telephone and Beyond. Pp.181. Hbk ISBN 0-7167-1311-X; pbk ISBN 0-7167-1336-5. (W.H. Freeman: 1982.) Hbk £11.80; pbk £5.50.

THOMPSON, J. (ed.). Power Sources 8, Research and Development in Non-Mechanical Electrical Power Sources. Proceedings of the 12th International Symposium held at Brighton, September 1980. Pp.630. ISBN 0-12-689155-9. (Academic: 1982.) £62.40, $128.

Computer Science

AGAJANIAN, A.H. Ion Implantation in Microelectronics: A Comprehensive Bibliography. Pp.255. ISBN 0-306-65198-X. (IFI/Plenum: 1981.) $85.

BOYER, R.S. and MOORE, R.J. (eds). The Correctness Problem in Computer Science. Pp.280. ISBN 0-12-122920-3. (Academic: 1981.) £14.20, $29.50.

HOCKNEY, R.W. and JESSHOPE, C.R. Parallel Computers. Architecture, Programming and Algorithms. Pp.416. ISBN 0-85274-422-6. (Adam Hilger, Bristol: 1981.) £22.50.

HOLDSWORTH, B. Digital Logic Design. Pp.338. Hbk ISBN 0-408-00404-5; pbk ISBN 0-408-00566-1. (Butterworths: 1981.) Hbk £17.95; pbk £9.95.

LEE, J.D. and LEE, T.D. Statistics and Computer Methods in BASIC. Pp.198. Hbk ISBN 0-442-30474-9; pbk ISBN 0-442-30475-7. (Van Nostrand Reinhold: 1982.) Hbk £11.50; pbk £5.50.

YOVITS, M.C. (ed.). Advances in Computers, Vol.20. Pp.283. ISBN 0-12-012120-4. (Academic: 1981.) $32.

Earth Sciences

BARAGAR, W. R. A. Geological Survey of Canada. Bulletin 314. Tectonic and Regional Relationships of the Seal Lake and Bruce River Magmatic Provinces. Pp.72. Pbk ISBN 0-660-10206-4. (Canadian Government Publishing Centre: 1981.). Pbk $10 (Canada) $12 (elsewhere).

BATES, R.L. and JACKSON, J.A. (eds). Glossary of Geology. 2nd Edn. Pp.751. ISBN 0-913312-15-0. (American Geological Institute, Virginia: 1980.) $60.

CHAGNON, S. A. Jr. (ed.). Metromex: A Review and Summary. Meteorological Monographs, Vol. 18, No. 40, October 1981. Pp.181. ISBN 0-933876-52-1. (American Meteorological Society: 1981.) Np.

DE MARSILY, G. and MERRIAM, D. F. (eds). Predictive Geology with Emphasis on Nuclear-Waste Disposal. Proceedings of Papers Presented at the 26th International Geological Congress in Paris, July 1980. Pp.206. ISBN 0-08-026246-5. (Pergamon: 1981.) £17.50, $35.

FRAZIER, J. W. (ed.). Applied Geography: Selected Perspectives. Pp.333. ISBN 0-13-040451-9. (Prentice-Hall: 1982.) £16.45.

GORDEY, S. P. Geological Survey of Canada. Bulletin 318. Stratigraphy, Structure and Tectonic Evolution of Southern Pelly Mountains in the Indigo Lake Area, Yukon Territory. Pp.44. Pbk ISBN 0-660-10620-5. (Canadian Government Publishing Centre, Ottawa: 1981.) Pbk $5 (Canada) $6 (elsewhere).

KANASEWICH, E.R. Time Sequence Analysis in Geophysics. 3rd Edn. Pp.480. ISBN 0-88864-074-9. (University of Alberta Press/Lea Valley Books, Hastings, E. Sussex: 1981.) $30.

NIO, S.-D., SHÜTTENHELM, R. T. E. and VAN WEERING, Tj. C. E. (eds). Holocene Marine Sedimentation in the North Sea Basin. Special Publication No. 5 of the International Association of Sedimentologists. Pp.515. Pbk ISBN 0-632-00858-X. (Blackwell Scientific: 1981.) Pbk £35.

SIMMONS, I. G. Biogeographical Processes. Processes in Physical Geography, Vol. 5. Pp.136. Pbk ISBN 0-04-574016-X. (George Allen & Unwin: 1982.) Pbk £4.25.

Biological Sciences

ALLEN, J. A. and SANDERS, H. L. Studies on the Deep Sea Protobranchia: The Subfamily Spinulinae (Family Nuculanidae). Bulletin of the Museum of Comparative Zoology, Vol. 150. No. 1, 12 January 1982. Pp.30. Pbk ISSN 0027-4100. (Harvard University: 1982.) Pbk $4.

ANDREWS, M.A. The Flight of the Condor. A Wildlife Exploration of the Andes. Pp.158. ISBN 0-00-219545-3. (Collins: 1982.) £12.95.

ARMS, K. and CAMP, P.S. Biology. 2nd Edn. Pp.942. ISBN 0-03-059961-X. (CBS College Publishing: 1982.) Np.

BAKER, R. R. Human Navigation and The Sixth Sense. Pp.138. ISBN 0-671-44390-9. (Simon & Schuster, New York: 1982.) $14.50.

BALL, I. R. and REYNOLDSON, T. B. British Planarians. Platyhelminthes: Tricladida. Keys and Notes for the Identification of the Species. Synopses of the British Fauna, 19. Pp.141. Hbk ISBN 0-521-23875-7; pbk ISBN 0-521-28272-1. (Cambridge University Press: 1982.) Hbk £16; pbk £6.95.

BARTH, R.H. and BROSHEARS, R.E. The Invertebrate World. Pp.646. ISBN 0-03-013276-2. (CBS Educational and Professional Publishing: 1982.) Np.

BEECHER, G.R. (ed.). Beltsville Symposia in Agricultural Research. 4, Human Nutrition Research. Papers presented at a Symposium held May 1979 at the Beltsville Agricultural Research Centre (BARC) Beltsville, Maryland. Pp.303. ISBN 0-916672-48-4. (Granada Publishing: 1982.) £20.

BEEKMAN, E.M. (ed. and translator). The Poison Tree. Selected Writings of Rumphius on the Natural History of the Indies. Pp.260. ISBN 0-87023-329-7. (The University of Massachusetts Press: 1981.) $20.

BELL, W. J. The Laboratory Cockroach. Experiments in Cockroach Anatomy, Physiology and Behaviour. Pp.161. Pbk ISBN 0-412-23990-6. (Chapman & Hall: 1981.) Pbk £6.95.

BELLAIRS, R., CURTIS, A. and DUNN, G. (eds). Cell Behaviour. A tribute to Michael Abercrombie. Pp.615. ISBN 0-521-24107-3. (Cambridge University Press: 1982.) £45.

BOGART, J. P. Chromosome Studies in *Sminthillus* from Cuba and *Eleutherodactylus* from Cuba and Puerto Rico (Anura: Leptodactylidae). Life Sciences Contributions, Royal Ontario Museum, No. 129. Pp.22. Pbk ISBN 0-88854-280-1. (Royal Ontario Museum, Toronto: 1981.) Pbk $2.

BOYLE, P.R. The Institute of Biology's Studies in Biology, No.134. Molluscs and Man. Pp.59. Pbk ISBN 0-7131-2824-0. (Edward Arnold, London: 1981.) Pbk £2.25.

BRIDGES, J.W. and CHASSEAUD, L.F. (eds). Progress in Drug Metabolism, Vol.6. Pp.321. ISBN 0-471-28023-2. (Wiley: 1981.) £24.75, $59.

BROUGHTON, W. J. (ed.). Nitrogen Fixation. Vol. 1, Ecology. Pp.306. ISBN 0-19-854540-1. (Oxford University Press: 1982.) £22.

BURGESS, R. L. and SHARPE, D. M. (eds). Forest Island Dynamics in Man-Dominated Landscapes. Ecological Studies. Anaylsis and Synthesis, Vol. 41. Pp.310. ISBN 3-540-90584-7. (Springer-Verlag: 1981.) DM73, $34.

BURTON, J. The Oxford Book of Insects, Pocket Edn. Pp.200. Pbk ISBN 0-19-217725-7. (Oxford University Press: 1982.) Pbk £2.50.

CERVÓS-NAVARRO, J. and FRITSCCHKA, E. (eds). Cerebral Microcirculation and Metabolism. Pp.500. ISBN 0-89004-590-9. (Raven: 1981.) $55.

COOMBS, J. and HALL, D. O. (eds). Techniques in Bioproductivity and Photosynthesis. Pp.171. Flexi ISBN 0-08-027383-1. (Pergamon: 1981.) Hbk np; flexi £7.50, $16.

COOPERSTEIN, S.J. and WATKINS, D. (eds). The Islets of Langerhans: Biochemistry, Physiology, and Pathology. Pp.497. ISBN 0-12-187820-1. (Academic: 1981.) $53.

CREASEY, W. A. Cancer: An Introduction. Pp.271. Hbk ISBN 0-19-502951-8; pbk ISBN 0-19-502952-6. (Oxford University Press: 1982.) Hbk np; pbk £10.

CSÁNYI, V. General Theory of Evolution. Studia Biologica Hungarica 18. Pp.123. Pbk ISBN 963-05-2823-1. (Akadémiaiai Kiadó, Hungarian Academy of Sciences, Budapest: 1982.) Pbk $7.

DAIBER, F.C. Animals of the Tidal Marsh. Pp.422. ISBN 0-442-24854-7. (Van Nostrand Reinhold: 1981.) £16.95.

DAUSSET, J. (ed.). Complexe Majeur d'Histocompatibilité de l'Homme. Pp.414. Pbk ISBN 2-257-10412-9. (Flammarion, Paris: 1981.) Pbk np.

DENOLIN, H. (ed.). Psychological Problems Before and After Myocardial Infarction. Advances in Cardiology, Vol. 29. Meeting, Mas d'Artigny, June 1981. Pp.156. ISBN 3-8055-3424-8. (Karger, Basel: 1982.) SwFr.98, DM117, $58.75.

DINNO, M.A. and CALLAHAN, A.B. (eds). Membrane Biophysics Structure and Function in Epithelia. Progress in Clinical and Biological Research Vol.73. A Symposium, Virginia Polytechnic Institute and State University, Blacksburg, May 1981. Pp.332. ISBN 0-8451-0073-4. (Alan R. Liss, New York: 1981.) £20.50, DM90.

DUKOR, P. *et al.* (eds). Cell Mediated Reactions Miscellaneous Topics. PAR Pseudo-allergic Reactions. Involvement of Drugs and Chemicals, Vol. 3. Pp.160. ISBN 3-8055-0960-X. (Karger, Basel: 1982.) SwFr.79, DM95, $47.50.

DUTRILLAUX, B. and COUTURIER, J. La Pratique de l'Analyse Chromosomique. Techniques de Laboratoire, 12. Pp.86. Pbk ISBN 2-225-74052-6. (Masson, Paris: 1981.) Pbk np.

FINK, W. L. and WEITZMAN, S. H. Relationships of the Stomiiform Fishes (Teleostei), with a Description of *Diplophos*. Bulletin of the Museum of Comparative Zoology, Vol. 150. No. 2, 12 January 1982. Pp.62. Pbk ISSN 0027-4100. (Harvard University: 1982.) Pbk $5.24.

FUJITA, T., TANAKA, K. and TOKUNAGA, J. SEM Atlas of Cells and Tissues. Pp.338. ISBN 0-89640-051-4. (Igahu-Shoin, Tokyo, New York: 1981.) $90, ¥17,000.

GARROD, D.R. and FELDMAN, J.D. (eds). Development in the Nervous System. The 5th Symposium of the British Society for Developmental Biology. Pp.403. ISBN 0-521-23493-X. (Cambridge University Press: 1982.) £40.

GORIO, A., MILLESI, H. and MINGRINO, S. (eds). Post-traumatic Periphal Nerve Regeneration: Experimental Basis and Clinical Implications. Pp.658. ISBN 0-89004-754-5. (Raven: 1981.) $50.

GOULD, J. L. Ethology: The Mechanisms and Evolution of Behaviour. Pp.544. ISBN 0-393-01488-6. (W. W. Norton: 1982.) $18.95.

GRESSER, I. (chief ed.) Interferon 1981, Vol.3. Pp.154. Pbk ISBN 0-12-302252-5. (Academic: 1981.) Pbk £9.20, $19.

GRESSITT, J. L. (ed.). Biogeography and Ecology of New Guinea, Vols 1 and 2. Monographiae, Vol. 42. Pp.982. ISBN 90-6193-094-4. (W. Junk, The Hague: 1982.) (Two vols) Np.

HABERMEHL, G. G. Venomous Animals and Their Toxins. Based on the Translation of the 2nd German Edn. "G. Habermehl, Gift-Tiere und ihre Waffen". Pp.195. Flexi ISBN 3-540-10780-0. (Springer-Verlag: 1981.) DM34, $15.90.

HARRIS, J. R. (ed.). Electron Microscopy of Proteins, Vol. 1. Pp.350. ISBN 0-12-327601-2. (Academic: 1981.) £35, $72.

IOACHIM, H. L. (ed.). Pathobiology Annual, Vol. 11, 1981. Pp.432. ISBN 0-89004-571-2. (Raven: 1981.) $42.50.

JOHN, P. C. L. (ed.). The Cell Cycle. Society for Experimental Biology Seminar Series, 10. Pp.276. Hbk ISBN 0-521-23912-5; pbk ISBN 0-521-28342-6. (Cambridge University Press: 1982.) Hbk £20; pbk £8.95.

JONES, D.G. The Institute of Biology's Studies in Biology, No.135. Neurones and Synapses. Pp.60. Pbk ISBN 0-7131-2825-9, (Edward Arnold, London: 1981.) Pbk £2.25.

KUHNEL, W. Pocket Atlas of Cytology and Microscopic Anatomy, 2nd Edn. Pp.289. ISBN 0-8151-5208-6/ 3-13-562402-1. (Year Book Medical Publishers/Georg Thieme Verlag: 1981.) £12.25.

KUNKEL, H.G. and DIXON, F.J. (eds). Advances in Immunology, Vol.31. Pp.324. ISBN 0-12-022431-3. (Academic: 1981.) $39.50.

KURSTAK, E. and KURSTAK, C. (eds). Comparative Diagnosis of Viral Diseases. Pp.429. ISBN 0-12-429703-X. (Academic: 1981.) $54.

LACKIE, J. M. and WILKINSON, P. C. (eds). Biology of the Chemotactic Response. Society for Experimental Biology Seminar Series, 12. Pp.177. Hbk ISBN 0-521-23305-4; pbk ISBN 0-521-29897-0. (Cambridge University Press: 1982.) Hbk £20; pbk £8.95.

LANGONE, J.J. and VAN VUNAKIS, H. (eds). Immunochemical Techniques, Part C. Methods in Enzymology, Vol.74. Pp.729. ISBN 0-12-181974-4. (Academic: 1981.) $58.

LASKER, R. (ed.) Marine Fish Larvae: Morphology, Ecology, and Relation to Fisheries. Pp.131. Pbk ISBN 0-295-95883-9. (University of Washington Press, Seattle & London: 1981.) Pbk $8.50.

LEISINGER, T. *et al.* (eds). Microbial Degradation of Xenobiotics and Recalcitrant Compounds. FEMS Symposium, No.12. Pp.416. ISBN 0-12-442920-3. (Academic: 1982.) £32, $66.

LI, C.H. (ed.). Hormonal Proteins and Peptides. Vol. X, ß-Endorphin. Pp.359. ISBN 0-12-447210-9. (Academic: 1981.) $44.50.

LLOYD, C.W. and REES, D.A. (eds). Cellular Controls in Differentiation. Based on the Unilever Jubilee Symposium held in Vlaardingen, Holland, December 1980. Pp.326. ISBN 0-12-453580-1. (Academic: 1981.) £12.40, $25.

LOBBAN, C. S. and WYNNE, M. J. (eds). The Biology of Seaweeds. Botanical Monographs, Vol.17. Pp.786. ISBN 0-632-00672-2. (Blackwell Scientific: 1981.) £45.

LOCK, M.A. and WILLIAMS, D.D. (eds). Perspectives in Running Water Ecology. Pp.430. ISBN 0-306-40898-8. (Plenum: 1981.) $49.50.

MAOLE, T. L. and HOFF, M. P. Gorilla Behavior. Van Nostrand Reinhold Primate Behavior and Development Series. Pp.290. ISBN 0-442-25152-1. (Van Nostrand Reinhold: 1982.) £27.20.

MASON, C. F. Biology of Freshwater Pollution. Pp.250. Pbk ISBN 0-582-45596-0. (Longman: 1982.) Pbk £5.50.

MATTHEWS, R.E.F. Plant Virology. 2nd Edn. Pp.897. ISBN 0-12-480560-4. (Academic: 1981.) $59.50.

McKAY, R., RAFF, M.C. and REICHARDT, L.F. (eds). Monoclonal Antibodies to Neural Antigens. Cold Srping Harbor Reports in the Neurosciences, Vol. 2. Pp.282. ISBN 0-87969-138-7. (Cold Spring Harbor Laboratory, New York: 1981.) $35 (US) $42 (elsewhere).

MESULAM, M.-M. (ed.). Tracing Neural Connections with Horseradish Peroxidase. Pp.251. Hbk ISBN 0-471-10028-5; pbk ISBN 0-471-10029-3. (Wiley: 1982.) Hbk £22, $52; pbk £11, $26.

MILLER, J. The Body in Question. Pp.352. Pbk ISBN 0-333-32978-3. (Papermac: 1982.) Pbk £5.95.

MILLER, P. C. (ed.). Resource Use by Chaparral and Matorral: A Comparison of Vegetation Function in Two Mediterranean Type Ecosystems. Ecological Studies: Analysis and Synthesis, Vol. 39. Pp.455. ISBN 3-540-90556. (Springer-Verlag: 1981.) DM98, $45.70.

MULLER, K.J., NICHOLLS, J.G. and STENT, G.S. (eds). Neurobiology of the Leech. Pp.320. ISBN 0-87969-146-8. (Cold Spring Harbor Laboratory, New York: 1981.) $36 (US) $43.20 (elsewhere).

NAHMIAS, A.J., DOWDLE, W.R. and SCHINAZI, R.F. (eds). The Human Herpesviruses: An Interdisciplinary Perspective. Proceedings of the International Conference, Emory University, Atlanta, Georgia, March 1980. Pp.721. ISBN 0-444-00553-6. (Elsevier/North-Holland: 1981.) $89, Dfl. 203.

NICHOLLS, D.G. Bioenergetics. An Introduction to the Chemiosmotic Theory. Pp.190. Hbk ISBN 0-12-518120-5; pbk ISBN 0-12-518120-5. (Academic: 1982.) Hbk £13.40, $27.50; pbk £5.50, $11.50.

NRIAGU, J.O. (ed.). Cadmium in the Environment Part II: Health Effects. Pp.908. ISBN 0-471-06455-6. (Wiley: 1981.) £85, $150.

PAOLETTI, R. and KRITCHEVSKY, D. (eds). Advances in Lipid Research, Vol. 18. Pp.320. ISBN 0-12-024918-9. (Academic: 1981.) $36.

PETERKEN, G.F. Woodland Conservation and Management. Pp.328. ISBN 0-412-12820-9. (Chapman & Hall: 1981.) £25.

PETO, R. and SCHNEIDERMAN, M. (eds). Quantification of Occupational Cancer, Banbury Report 9. Pp.750. ISBN 0-87969-208-1. (Cold Spring Harbor Laboratory: 1981.) $89 (US) $106.80 (elsewhere).

PRASAD, K.N. and VERNADAKIS, A. (eds). Mechanisms of Actions of Neurotoxic Substances. Pp.235. ISBN 0-89004-638-7. (Raven: 1982.) $26.50.

RICHARDS, D. The Topic of Cancer. When the Killing Has to Stop. Pp.147. ISBN 0-08-025937-5. (Pergamon: 1982.) £4.95.

SEN, D.N. and JHA, P.K. (eds). Indian Review of Life Sciences, Vol. 1, 1981. Pp.273. (no ISBN). (University of Jodhpur, Jodhpur-342 001, India: 1981.) Np.

SERET, B. Poissons de Mer de L'Ouest Africain Tropical. Initiations-Documentations Techniques, No. 49. Pp.416. Pbk ISBN 2-7099-0600-7. (ORSTOM, Paris: 1981.) Np.

SKRABANEK, P. and WALSH, A. (eds). Bladder Cancer. A Series of Workshops on the Biology of Human Cancer, Report No. 13. UICC Technical Report Series, Vol. 60. Pp.192. Pbk ISBN 92-9018-060-9. (International Union Against Cancer, Geneva: 1981.) Pbk SwFr. 32, $16.

SKULACHEV, V.P. and HINKLE, P.C. (eds). Chemiosmotic Proton Circuits in Biological Membranes: In Honor of Peter Mitchell. Pp.633. ISBN 0-201-07398-6. (Addison-Wesley: 1981.) $29.50.

SNELL, K. (ed.). Developmental Toxicology. Pp.350. ISBN 0-7099-2306-6. (Croom Helm, London: 1981.) £22.50.

STRÖM, I. Laboratory Animal Science: A Selected Bibliography. Acta Physiologica Scandinavica, Supplement 498. From the Library of Uppsala University and Swedish University of Agricultural Sciences. Pp.38. (no ISBN) (Uppsala University, Sweden: 1981.) Pbk np.

TIVY, J. Biogeography; A Study of Plants in the Ecosphere. 2nd Edn. Pp.459. Pbk ISBN 0-582-30009-6. (Longman; 1982.) Pbk £6.95.

TOWE, A.L. and LUSCHEI, E.S. Handbook of Behavioral Neurobiology. Vol. 5, Motor Coordination. Pp.640. ISBN 0-306-40613-6. (Plenum: 1981.) $45.

WELTY, J.C. The Life of Birds. 3rd Edn. Pp.754. ISBN 0-03-057917-1. (CBS Educational and Professional Publishing, New Jersey: 1982.) Np.

WOBESER, G. A. Diseases of Wild Waterfowl. Pp.300. ISBN 0-306-40746 -7. (Plenum: 1981.) $39.50.

WOOLSEY, C. N. (ed.). Cortical Sensory Organization. Vol. 1, Multiple Somatic Areas. Pp.245. ISBN 0-89603-030-X. (Humana Press, New Jersey: 1981.) $34.50 (US) $44.50 (elsewhere).

Applied Biological Sciences

BAZÁN, N. G., PAOLETTI, R. and IACONO, J. M. (eds). New Trends in Nutrition, Lipid Research, and Cardiovascular Diseases. Current Topics in Nutrition and Disease, Vol. 5. Pp.314. ISBN 0-8451-1604-5. (Alan R. Liss, New York: 1981.) £20.50, DM90.

BELLINA, J. H. (ed.7. Gynecologic Laser Surgery. Based on the Proceedings of the International Congress held January 1980 in New Orleans, Louisiana. Pp.480. ISBN 0-306-40741-8. (Plenum: 1981.) $45.

CUNNINGHAM, D. *et al.* (eds). Control of Cellular Division and Development. Part A, Progress in Clinical and Biological Research, Vol. 66A & 66B. Proceedings of the ICN–UCLA Symposium held at Keystone, Colorado, March 1980. Vol. 66A pp.636, ISBN 0-8451-0156-0; Vol. 66B pp.458, ISBN 0-8451-0157-9. (Alan R. Liss, New York: 1981.) Vol. 66A £48.90, DM125; Vol. 66B £40.70, DM179.

GOLDMAN, L. (ed.). The Biomedical Laser: Technology and Clinical Applications. Pp.342. ISBN 3-540-90571-5. (Springer–Verlag: 1981.) DM98, $41.70.

INTERNATIONAL AGENCY FOR RESEARCH ON CANCER. IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans. Some Antineoplastic and Immunosuppressive Agents, Vol. 26. IARC Working Group which met in Lyon, October 1980. Pp.294. Flexi ISBN 92-832-1309-2. (WHO: 1981.) SwFr.41.

ISRAEL, Y. *et al.* (eds). Research Advances in Alcohol and Drug Problems. Vol. 6, Research Advances in Alcohol and Drug Problems Series. Pp.521. ISBN 0-306-40672-1. (Plenum: 1981.) $59.50.

NORRIS, J. R. and RICHMOND, M. H. (eds). Essays in Applied Microbiology. Pp.352. ISBN 0-471-27998-6 (Wiley: 1981.) Hbk £19.50, $46; pbk £4.95, $11.95. This book is available in separate chapters in sets of five.

OWEN, C. A. Jr. Wilson's Disease: The Etiology, Clinical Aspects, and Treatment of Inherited Copper Toxicosis. Pp.215. ISBN 0-8155-0879-4. (Noyes Publications, New Jersey: 1981.) $28.

ROSE, A. H. (ed.). Microbial Biodeterioration. Economic Microbiology, Vol. 6. Pp.516. ISBN 0-12-596556-7. (Academic: 1981.) £38.60, $79.50.

SMITH, J. E. Biotechnology. The Institute of Biology's Studies in Biology, No.136. Pp.73. Pbk ISBN 0-7131-2835-6. (Edward Arnold, London: 1981.) Pbk £2.60.

STONEHOUSE, B. Biological Husbandry. A Scientific Approach to Organic Farming. Pp.352. ISBN 0-408-10726-X. (Butterworths: 1981.) £25.

STOWARD, P. J. and POLAK, J. M. (eds). Histochemistry: The Widening Horizons of its Applications in the Biomedical Sciences. Pp.293. ISBN 0-471-10010-2. (Wiley: 1981.) £25, $60.

TOURAINE, J. L. *et al.* (eds). Transplantation and Clinical Immunology. Vol. 13, Proc. of the 13th International Course, Lyon, June 1981. Pp.273. ISBN 90-219-0487-X. (Excerpta Medica, Amsterdam: 1981.) $55.75, Dfl.120.

WILKERSON, R. D. (ed.). Cardiac Pharmacology. Physiologic and Pharmacologic Bases of Drug Therapy, 2. Pp.447. ISBN 0-12-752050-3. (Academic: 1981.) $52.

WOLF, S. and MURRAY, A. K. Composition and Function of Cell Membranes. Application to the Pathophysiology of Muscle Diseases. An edited transcript of the Totts Gap Colloquium held May 1980. Pp.287. ISBN 0-306-40883-X. (Plenum: 1981.) $35.

Psychology

IMMELMANN, K. *et al.* (eds). Behavioral Development. The Bielefeld Interdisciplinary Project. Pp.754. Hbk ISBN 0-521-24058-1; pbk ISBN 0-521-28410-4. (Cambridge University Press: 1982.) Hbk £35; pbk £12.50.

LINDEMANN, J. E. Psychological and Behavioral Aspects of Physical Disability. A Manual for Health Practitioners. Pp.426. ISBN 0-306-40776-0. (Plenum: 1981.) $24.50.

MENDLEWICZ, J., COPPEN, A. and VAN PRAAG, H. M. (eds). Depressive Illness. Biological and Psychopharmacological Issues. Advances in Biological Psychiatry, Vol. 7. Symposium, Amsterdam, September 1980. Pp.240. Pbk ISBN 3-8055-2482-X. (Karger, Basel: 1981.) SwFr.74, DM89, $44.50.

General

ALTMAN, I., WOHWILL, J. F. and EVERETT, P. B. Transportation and Behaviour. Human Behaviour and Environment: Advances in Theory and Research, Vol. 5. Pp.285. ISBN 0-306-40773-6. (Plenum: 1981.) $25.

BARTELS, A. Garten Geholze. Pp.496. (no ISBN) (Verlag Eugen Ulmer: 1981.) DM168.

BERMAN, M. The Reenchantment of the World. Pp.357. Hbk ISBN 0-8014-1347-0; pbk ISBN 0-8014-9225-4. (Cornell University Press: 1982.) Hbk £24.25, $43.10; pbk £5.95, $11.10.

BUTLER, P. H. Point Group Symmetry Applications. Methods and Tables. Pp.567. ISBN 0-306-40523-7. (Plenum: 1981.) $55.

CAPLAN, A. L. and CALLAHAN, D. (eds). Ethics in Hard Times. Pp.296. ISBN 0-306-40790-6. (Plenum: 1981.) $25.

COX, K. R. and GOLLEDGE, R. G. (eds). Behavioural Problems in Geography Revisited. Pp.290. Hbk ISBN 0-416-72430-2; pbk ISBN 0-416-72440-X. (Methuen: 1982.) Hbk £12.50; pbk £5.95.

DELSON, R. M. Readings in Caribbean History and Economics: An Introduction to the Region. Caribbean Studies, Vol. 1. Pp.336. ISBN 0-677-05280-4. (Gordon & Breach: 1981.) Hbk $59.50, pbk np.

EISENREICH, S. J. (ed.). Atmospheric Pollutants In Natural Waters. Pp.512. ISBN 0-250-40369-2. (Butterworths: 1981.) £23.10.

ENNS, R. H. *et al.* (eds). Nonlinear Phenomena in Physics and Biology. Proceedings of a NATO Advanced Study Institute held August 1980, at the Banff Centre, Alberta, Canada. Pp.609. ISBN 0-306-40880-5. (Plenum: 1981.) $75.

EVAN, W. M. (ed.). Knowledge and Power in a Global Society. Sage Focus Editions, Vol. 30. Pp.320. Hbk ISBN 0-8039-1659-0; pbk ISBN 0-8039-1660-4. (Sage, California: 1981.) Hbk $20;pbk $9.95.

GEHLBACH, F. R. Mountain Islands and Desert Seas: A Natural History of the US – Mexican Borderlands. Pp.298. ISBN 0-89096-118-2. (Texas A&M University Press: 1982.) $19.95.

HAITH, D. A. Environmental Systems Optimization. Pp.306. ISBN 0-471-08287-2. (Wiley: 1982.) £18.55, $33.

HEATH, A. F. (ed.). Scientific Explanation. Papers based on Herbert Spencer Lectures given in the University of Oxford. Pp.123. ISBN 0-19-858214-5. (Oxford University Press: 1981.) £7.95.

nature

13 May 1982

Will there be war in space?

The sheer scale of the Soviet space programme is awesome. Has the time come when the Soviet government must say what it hopes to achieve?

In June the name of Jean-Loup Chrétien will become a household word in the Soviet Union and France. Chrétien will be the first Western astronaut to be whooshed into space by a Soviet rocket and to ride in a *Soyuz* spacecraft, visit a Soviet space station briefly and, if all goes well, return to Earth. He will join the list of Polish, Czech, East German, even Cuban, cosmonauts whom the Soviet Union has made guests in space, and who have been given a first-hand look at the civilian side of that technology about which the Soviet Union is most sensitive and proud: manned space flight. At about the same time, in Washington, where the Soviet space programme generally draws a big yawn, the President will be in the final stages of a critical decision on whether to proceed with a US manned space station to serve the United States and Western Europe.

The Soviet space programme has received little attention in Europe or the United States. Yet its objectives and momentum will influence East–West relations and the arms race in the 1980s and beyond. True, the Western media are delighted to talk up "Star Wars" fantasies, the US shuttle, Ariane or Soviet space weapons. But they have overlooked the main substance of the Soviet programme, lucidly outlined in a recent book by NASA watcher of the Soviet skies, James E. Oberg*. Since the early 1970s, the Soviet manned space programme, which had been in a shambles after the death of its early architect Sergey Korolev in 1966, has concentrated on a single goal: *Kosmograd*. This is the code word for "Space Colony Number 1, made in the Soviet Union". In an assessment shared by many observers of the Soviet manned space programme, Oberg describes how the Soviet Union is amassing the building blocks of this capability.

First there was *Soyuz*, the free-flying craft in which cosmonauts journey up and down from space and make (brief?) orbital stays for scientific work or military reconnaissance. These flights started in 1967 and numbered 40 by May 1981. Then came the *Salyut* space stations, the first one launched in 1971. These have been orbiting space homes for repeated and successively longer visits by cosmonauts. The world's record of 185 days was reached aboard *Salyut 6* in 1980. One of the astronauts who performed the feat had just spent another 175 days in space, and so has had nearly a year of weightlessness. Finally came *Progress*, the unmanned robot freighter tanker craft described in Moscow as a "reliable cargo barge to ensure year-long operation by orbital stations". *Progress* craft bring lonely cosmonauts fuel, food, fresh water, medicines, horseradish, newspapers, mail and other things. By now there is extensive experience in sending up these vehicles, docking the *Soyuz* and *Progress* craft with the *Salyut*, moving them from one *Salyut* port to another to admit a second visiting craft and making repairs in space, including emergency repairs to the exterior of the craft.

So engineering experience is accumulating, but *Kosmograd* also requires mastery of space biology and medicine as well as experience with prolonged human confinement. The goal of sustaining life in space for a year or more is within sight. Although the large green cucumber that the crew on one *Salyut* showed to astonished mission controllers on television proved to be a joke, with practice the Soviets have managed to get seeds growing in space. Space gardens could provide 20 to 40 per cent of the bulk food consumed in orbit, but it remains a mystery why space-grown plants have not yet grown flowers or seeds. Water aboard the *Salyut* is now 50 per cent recycled, and the addition of a urine still could achieve 90 per cent recycling. Banks of *Chlorella* algae have been used to absorb carbon dioxide and give off fresh oxygen: 80 per cent recycling has been achieved on this front. But things have not always gone well. Cosmonauts set to achieve a new endurance record aboard *Salyut 4* pleaded early on to come home because a green mould was spreading all over the cabin. Mission control ignored these pleas until the endurance record had been broken (61 days).

The objective of all this is a permanent orbiting station. *Salyut*, housing two, will be succeeded soon by a larger spaceship housing twelve. According to Oberg, these craft could even be sent to the Moon with the help of the huge proton booster (whose shadowy cousin, the superbooster, apparently failed so often in the late 1960s that it prevented a photofinish in the race to the Moon against the United States).

But such a feat will, on past experience, require a high price in human life. Cosmonauts may be heroes in the Soviet Union, but few seem to live to enjoy it. Recovery teams have often arrived at a returned capsule dangling by its parachute lines in a tree or sitting in some swamp to find the entire crew dead. There was even an unauthorized space-walk in which the cosmonaut's life support cord was unattached and when his life was saved only by the quick thinking of his colleague. Yuri Gagarin, whose 108-minute flight in 1961 amazed the world, became a national hero but he was fated to eulogize comrade after comrade at one funeral after another. (His own life was thrown away cheaply in an aircraft crash.) Even Korolev's death on an operating table, which cost the Soviet space programme its single most important figure, seems to have been unnecessary. There have also been close shaves with death, as when the cosmonauts who missed their landing area by 2,000 miles had to spend their first night back on Earth in the Siberian snows, hiding in their spacecraft from the wolves that snapped outside.

What is the underlying purpose of all this activity? There are obvious civil benefits, in telecommunications for example. And by now, both the United States and the Soviet Union take for granted the stabilizing effects of each other's use of satellites for surveillance. Even so, there is growing concern in the United States at what seems to be the huge scale of the military part of the Soviet space programme. Oberg says there are two "star cities", one for cosmonauts and their guests and the other for the military. A recent study by Marcia Smith of the Congressional Research Service concludes that as much as 70 per cent of the Soviet effort may be exclusively military, while nobody has yet contradicted the assertion by Richard de Lauer, chief scientist at the Pentagon, that spaceborne Soviet lasers could threaten US spacecraft as early as 1983. Technical considerations make the speculation unconvincing; for the time being conventional rocket-driven explosives are probably a better bet. But as in other aspects of the relationship between the United States and the Soviet Union, policy may be determined by one state's perceptions of what the other is about. The fact that the Soviet Union successfully launched 98 satellites last year, compared with a mere 18 from the United States, is certain to carry weight with the White House.

What will be the outcome? Since the United States and the

*Oberg, J.E. *Red Star in Orbit* (Harrap, London and Random House, New York).

0028-0836/82/190091-02$01.00

Soviet Union ratified the Outer Space Treaty in 1967, they have been bound not to put nuclear weapons or other "weapons of mass destruction" into orbits about the Earth. The treaty also requires them not to establish military bases on celestial bodies. Would that injunction apply to a space station occupied exclusively by the military? Or to a satellite equipped to knock out other people's satellites but which contained no people, only instruments? The Soviet Union should come clean about the objectives of its massive programme. The terrestrial arms race is uncomfortable enough. The prospect of a race to put flesh on the bones of science fiction is alarming. Meanwhile, the odds are shortening that President Ronald Reagan, due to greet the next return of the shuttle *Columbia* at Edwards Air Force Base in California on 2 July, will not merely proclaim that the shuttle is a going concern but that the United States will embark on the next logical step, the building of a space station. As Kennedy did before him, Reagan may find that the Soviet Union's activities justify an expanded programme in the United States.

Academic suicide

The University f London now has a plan for its future; but will it survive that long?

The University of London seems well on the way to making an even greater hash of its affairs than seemed likely a few months ago (see *Nature* 14 January, p.86). For the past year, it has been clear that the consequences of the British government's reduction of support for universities, the limits imposed by the University Grants Committee on the number of students to be allowed at individual universities and the effect on the income of universities of frightening away students from overseas would be especially damaging for the largest British university and, because of its federal constitution, administratively the most complicated. But none of that can excuse the way in which the university has let time slip through its fingers until last week when, with the deadline approaching for deciding how it should finance itself next year, the university's senate was jockeyed into accepting a plan for the future whose analytical foundations are, to say the least of it, shaky and whose academic implications are unknown. The only virtue in the way in which the university has managed its affairs is that this sorry tale may serve as a warning to other academic institutions.

What has happened in the University of London is a kind of pantomime. Two years ago, the then vice-chancellor, seeing the way the wind was blowing, set up a Committee on Academic Organization under an independent chairman. The plan was to recommend a strategy for the university as a whole. Halfway through this undertaking, however, a new vice-chancellor gave the committee narrower terms of reference and set up six other committees, again with independent chairmen, to suggest how the pattern of teaching in different fields of study should be reorganized. (One committee went off on the wrong tack, preferring terms of reference other than those given to it; mercifully, with a different chairman; a set of recommendations has been cobbled together just in time.) Inevitably, these subject committees have produced proposals for moving the teaching of various subjects from one place to another in the federated university which are inevitably in conflict with each other, so yet another committee — called the Joint Planning Committee — was asked to make sense of them and of the volume of comment and special pleading with which the university was deluged. That committee concludes, in its report, that the recommendations of the six subject committees were often based "on inadequate information" (ill-informed) or "cannot be justified in the light of the facts" (or in simple language are wrong). Searching, nevertheless, for some straw at which to clutch, the committee seems to have seized on one of the recurrent themes in the recommendations of the subsidiary committees — physics, or classics, or economics, should be taught at fewer sites. So why not simply concentrate the whole of the university, now spread between a dozen non-medical colleges, on a smaller number of places, say five or six? This is the proposal that the senate of the university found itself having to accept last week.

The most curious feature of the affair is that this radical decision has been reached with hardly any discussion in the university as a whole of the academic implications, let alone the implications for those who work in and for the university. And what is now intended will create further internal contradictions within what is already a sufficiently cumbersome academic machine. The university grew to its present size like Topsy, by the enthusiastic accretion of almost any college, large or small, willing to join the fold. As a result, it now consists of an exceedingly inhomogeneous collection of establishments, some large and some small, some academically excellent, some less so. Two are for practical purposes specialized in the scientific fields. One (Birkbeck College) specializes in part-time students. One small college is not in London at all but in Surrey, while there are agricultural and veterinary colleges in Kent and Hertfordshire respectively. The tidy administration of such a motley collection of establishments is acknowledged to be impossible. Since the late 1960s, the university has chosen to make a virtue out of diversity, letting its separate parts run themselves without much reference to the centre. Now, by a simple show of hands, that policy has been turned on its head. The committee whose recommendations will shape the future of the university pays the usual lip-service to diversity before recommending that it should be done away with.

The university's problem is uniquely difficult — which is all the more reason why it should have been dealt with intelligently. Given that the pressure on the university's finances is likely to continue more or less indefinitely, a measure of concentration is clearly in the long run prudent. Some of the London colleges, seeing the way the wind is blowing, are already making mergers among themselves. The merger (or disappearance) of the rest will now no doubt be forced by the decisions taken centrally by the university on the distribution of funds among its several parts in the years immediately ahead. In the process, it is inevitable that much damage will be done. Many of the shotgun marriages the university is now committed to arranging will be marriages of incompatibles. Undergraduate teaching will become more specialized, and more uniform. The pattern of research will be changed in an arbitrary fashion. And the end result will be a federation of colleges so much more like each other that the advantages of federation will melt away. If the arrangements now agreed are pushed through, it cannot be long before the various parts of the University of London go their separate ways. Is that what is intended?

The peculiar characteristic of the damage that academics repeatedly inflict upon themselves is that it is always done at the eleventh hour (and sometimes the thirteenth). In London, an acceptable solution may already be beyond reach. But that possibility does not mean that it is pointless to look for a better solution. A few clarifying assumptions are obvious starting points. The Imperial College of Science and Technology, in all but name a university in its own right, is rightly acknowledged to be too important to be enmeshed in the financial administration of the University of London (whence the convention that it deals separately with the University Grants committee); it should now be invited to go its own way. The London School of Economics, smaller but equally distinguished and important, should be given some spell of time, say ten years, to grow at the expense of the parts of the university now under threat and then also politely booted out. The rump would be three kinds of colleges — large or largeish place such as University College, smaller colleges concerned principally with the education of undergraduates, and a rag-bag of specialist institutions training veterinarians or assisting with research in outlandish languages. There is ample evidence that each kind of place is valued; what makes their administration impossible is that they are all lumped together. So why not replace them with three (or even more) different universities? When the University of Paris, as now reconstituted, consists of thirteen separate institutions, is it necessary that the largest city in Britain should be condemned to having only one?

0028-0836/82/190092-01$01.00

Polish university compromise

Buffer between academics and government

Poland's new Higher Education Act, the subject of eighteen months' negotatiations and confrontations — including the threat, last September, of a nationwide university strike — became law last week. The act makes several formal concessions to academics' aspirations for self-governance, but at the same time reserves to the Minister of Science, Higher Education and Technology the last word on a number of key issues.

Students, likewise, get a compromise deal — they have won the right to express their views and form associations but will not be allowed to stage protest actions since, it was explained during the parliamentary debate on the bill, "they do not belong to the community of employees".

The new act gives the minister power to monitor the resolutions adopted by university self-governance bodies and decisions taken by university rectors, and to reverse them should he decide they are inconsistent with the law. Rectors of universities and other higher colleges will be elected by the bodies themselves, the electors being either the university senate or a special collegium. The minister, however, retains the right of veto over elections — providing that he registers his objection within fourteen days.

On these and other questions, there is to be a Main Council for Science and Higher Education to hold the ring, and whose role has been negotiated between the ministry and the Social Drafting Commission — a body representing academics, students, scholars, the trade unions and the Church, which at one stage proposed its own draft bill. The council is in effect a revival under a new name of a moribund body meant to act as an intermediary between academics and the ministry.

The new act gives the council the status of the supreme elective organ of the academic community, with powers to decide on the direction of research and the organization of studies.

The Social Drafting Commission, however, had sought even wider powers for the Main Council. In the event, as Professor Zbigniew Resich, the head of the commission, said last week, "a whole series of important powers" had passed into the minister's hands.

Another major issue has been the dismissal of academic staff. The Social Drafting Commission, said Professor Resich, had considered that a lecturer's contract could be terminated only after disciplinary proceedings. In its final form, the act now contains other possibilities, including the dismissal of junior members of staff (up to and including assistant professors) after two "negative assessment reports". It is not yet clear who would make these assessments, although it seems that the rights of the professorial body will be safeguarded, since any such dismissal must have the consent of the Council of State. But there is some evidence that "assessments" by bodies outside the university are envisaged, another limitation of "self-governance".

Similarly, although the act formally acknowledges the principles of academic freedom, including differences of "world outlook", implementation of these principles, in undergraduate teaching at least, will be constrained by what Professor Resich called "strong emphasis" in the act on the "need to teach and bring up youth in accordance with socialist principles".

These other features of the act are intended to emphasis the state character of higher education in Poland, according to the minister, Dr Benon Miskiewicz. He himself was "following with anxiety" the participation of students and academics in illegal demonstrations, and stressed the role of the universities in ideological and political education.

His concern was timely; in Warsaw alone, six higher education establishments were represented in the arrests following last week's demonstratrations. Students were equally prominent in the demonstrations in Szczecin Gliwice and in the Gdansk-Sopot-Gdynia conurbation.

These demonstrations, which took place in spite of the threat of heavy summary sentences under the emergency regulations, clearly reflect the students' bitterness at the banning of the Independent Students' Association (NZS) and the loss of many of the privileges granted in 1981. The new Higher Education Act is hardly likely to assuage their resentment. **Vera Rich**

Contraction for London university

A proposal by a Joint Planning Committee that the teaching of science in the University of London should be concentrated on five sites was accepted by the senate of the university at its meeting on 29 April. No timetable for the concentration of the university is provided.

This is the university's response to the prospect of an income that will have shrunk by about 15 per cent between 1981-82 and 1983-84, and to earlier reports by the six committees set up to recommend rationalization of teaching in six subject areas. Most committees recommended that teaching should be concentrated at fewer sites.

The planning committee, however, says that these reports reflect "inadequate information" and make comments "that cannot be justified in the light of the facts". In practice, the committee responsible for biological sciences, under its independent chairman Sir James Beament, began last October by assuming that student numbers need not fall as sharply as had been decreed. The committee's second attempt under a second chairman follows the general pattern of recommending concentration on five sites.

The joint planning committee itself is hoping for some relief from the rigours prescribed by the University Grants Committee (UGC). When added together, the recommendations of the six subject committees on student numbers for the year 1983-84 imply that student numbers would then be 1,000 greater than the UGC target. The planning report says that the grants committee has been asked to agree to a larger target number or, alternatively, that the university should have an extra year in which to make the adjustment.

On research support, the planning committee applauds the study now being carried out by the court of the university of the suggestion that financial allocations to the separate London colleges should deal separately with teaching and research. This proposal, put forward last November by Sir James Lighthill, Principal of University College, is regarded as a way of helping to avoid the consequences of the erosion of the dual-support system in the past few years, and also of making explicit the reasons why some colleges in London appear to be dealt with more generously than others. But it seems unlikely that anything will come of this before the court of the university shares out the funds available for the next academic year.

The planning committee's recommendations are that undergraduate science teaching in the university should be concentrated at five sites, where all science subjects would be taught. There are similar proposals for concentration in the arts and the humanities.

The next step in the saga of the University of London will come on 19 May, when the court of the university is to meet for a preliminary assessment of how to divide next year's grant. The exact amount will not be known until the university has heard from UGC at the end of May.

Opinion among senior members of the university is divided about the extent to which the new policy will influence next year's distribution of funds within the university. Some think that the court will be bound to distribute what money there is as the planning committee says it should. Others hold that the university will have to respect the autonomy of its colleges. ●

0028-0836/82/190093-01$01.00

Paris computer centre

Orsay protests

The *Université de Paris-Sud* at Orsay, one of the principal centres of physics in France, has withdrawn from a plan to assist in the installation and development of the "largest" commercial computer in the world, a Cray-1. The computer is to be installed on the new site of the *Ecole Polytechnique* just a mile or two from the university, but Paris-Sud will have nothing to do with it (except to make a little use of it, once it is running). Orsay physicists have demonstrated over the matter and, at the same time, the scientific director of Paris-Sud's own computer laboratory, *Paris-Sud Informatique* (PSI), has threatened to resign on 20 June unless the government agrees to keep his own laboratory going.

The underlying issue is tangled, and the tangle goes right to the top. *Paris-Sud's* president, physicist Roland Omnès, is going to see the minister of education, Alain Savary, about the matter. But a remark that funding for PSI might not be renewed came from the ministry of research and technology, which is attempting to establish a computer policy for universities and research centres; and the two ministers, Savary and the research minister Jean-Pierre Chevènement, are not the best of friends.

The ministry of defence is also involved, both through its control of the *Ecole Polytechnique* and through the long-expressed wish of the military to purchase a Cray-1 (reportedly vetoed by Washington). So is the ministry of industry, which is trying to revitalize the French computer industry. Given the ever-present IBM lobby and the traditional hostility between the *grandes écoles* (such as the *Ecole Polytechnique*) and the universities, the result is a highly political problem.

At root, however, the dispute is technical. Cray-1 is a "vectorial" computer, whose great power depends on the parallel processing of a series of numbers. It is thus stretched for linear algebra (matrices and so on) only where the sizes of arrays are very large. Physicists generally do not need such arrays (except for modelling large systems such as the atmosphere) and for them the Cray-1 is far from being the most cost-effective machine. So when PSI computer engineers went to the *Ecole Polytechnique* to collaborate over Cray-1 development, they found themselves in conflict with *polytechnique* interests — which were, largely, in engineering and systems modelling. This led to personal conflict and ultimately Omnès withdrew the PSI team from the collaboration.

Since then the atmosphere has become even more heated, with PSI staff claiming, in effect, that the *Ecole Polytechnique* staff are in the hands of IBM; and indeed, the *polytechnique* has scrapped plans to link the Cray-1 to a French-built CII-Honeywell-Bull EBS-8 (a link PSI was to work on) and will go for an IBM instead. The final crunch was a comment made at a meeting by a ministry of research and technology official, who said that Chevènement would probably not sanction the renewal of PSI equipment (a ten-year-old UNIVAC), and that Orsay physicists should use the Cray-1 "only three kilometres away".

Omnès, however, is calm. He believes official thinking is moving his university's way, and that this is not the time for a big public campaign. Nevertheless, some of his more vocal staff disagree with him. Nor does time seem to be on his side. A ministry of research spokesman said this week that it did not want to replace university computers in response to "piecemeal" pressure, and that it would take "several months" to establish a coherent policy. In the meantime the (part-time) scientific director of the Orsay computer centre may have resigned and some of his full-time technical staff, uncertain about their future, may have begun to look for other jobs. That — ultimately — would leave one of France's leading laboratories without a computer.

Robert Walgate

Schering-Plough take $30 million plunge

Palo Alto

DNAX Ltd, a small Californian company best known for having some pioneers in genetic engineering on its scientific advisory board, seems likely to be acquired by the New Jersey pharmaceutical company Schering-Plough for $29.4 million. Schering-Plough already owns 13 per cent of another leading biotechnology company, Biogen SA.

Schering-Plough is to pay $19 million in cash and is offering 340,000 shares of its common stock valued at approximately $10.5 million. The common stock is to be distributed over 5 years under an incentive plan to the 18-member advisory board and staff of 40 scientists. Preferred shares of DNAX are at present owned by a Swiss group and a US venture capital unit of Elf Aquitaine, the French oil company.

Founded in 1980, DNAX has not so far launched any products or filed any patents. Its scientific advisory board boasts three Nobel laureates: Paul Berg and Arthur Kornberg, both of Stanford University, and George Palade of Yale University.

Alejandro Zaffaroni, DNAX's founder, hopes that the company will eventually manufacture the protein antibodies made by the human body when it is combating infections and inflammatory diseases and allergies, and hopes to have specific macromolecules ready for testing in four to five years.

Zaffaroni is also founder and president of Alza, a company specializing in the delivery of drugs to target organs within the body. Alza's common shares in DNAX, valued at $5.3 million, will be bought by Schering-Plough in its acquisition of DNAX.

In his search for a suitable buyer for DNAX, Zaffaroni says he approached more than 20 companies in Europe, Japan and the United States. Schering-Plough's purchase of DNAX will be the first instance of a major pharmaceuticals company acquiring an entire biotechnology firm.

Charlotte K. Beyers

Law of the Sea

Soviet enterprise

In advance of the inconclusive end to the UN conference on the law of the sea, the Soviet Union last month passed enabling legislation to cover the granting of prospecting and exploitation licences to Soviet enterprises wishing to engage in undersea mining beyond the continental shelf.

The decree of the Supreme Soviet now published describes the measures as "temporary" and stresses that the Soviet Union "is and will continue to be in favour of the settlement of urgent problems of the legal regime of the world ocean on an international basis . . . taking into account the legitimate interests of all states". In the absence of any such legislation, however, the Soviet Union has been "compelled" to protect its interests by granting off-shelf licences to Soviet mineralogical concerns.

The decree was clearly an answer to Western proposals on off-shelf mining. In particular, the head of the Soviet delegation to the law of the sea conference, deputy foreign minister S. P. Kozyrev, categorically rejected Western proposals to grant licences on a "first-come first served" basis. This, he said, was a principle "widely employed during the colonial enslavement of countries and peoples".

Commenting on the new decree, the daily *Socialisticheskaya Industriya* stressed the interest of US companies in off-shelf strategic minerals which, it alleged, had pressurized President Reagan to renege on earlier understandings. The new decree has therefore been interpreted by some commentators as simply a tactical move to strengthen the Soviet position at the bargaining table.

The provisions of the decree, however, are considerably more detailed than would be necessary for such a purpose. It goes into considerable detail about safety zones around deep drilling installations, the notification of temporary navigation hazards, disputes over rights with the "agencies of foreign states" and the setting up of a special fund under the Council of Ministers to which Soviet licence-holders must pay part of their profits.

0028-0836/82/190094-02$01.00

Moreover, a time-lag has been built into the decree, presumably to allow the Soviet Union to acquire the necessary technology. No off-shelf operations, it states, can begin before 1 January 1988.

Several clauses deal with possible non-Soviet participation in off-shelf operations. This would presumably be of two types — with developed and with developing countries. One likely partner in the first category is East Germany, with which the Soviet Union (together with Poland) is already linked in the Baltic drilling consortium *Petrobaltyk* to which the Germans have made a considerable technological contribution. The decree, however, stresses the Soviet Union's role in "assisting" its future partner in the development of technology and the training of personnel, suggesting participation by developing countries. (This would fit in with Kozyrev's rejection of the "colonial" overtones of the Western proposals.)

The developing country would in the first instance presumably provide a shore base and other facilities. Any minerals recovered, according to the decree, would become Soviet property, though in some cases, some or all of them could revert to the partner state. The concept of the Soviet Union building up Third-World off-shelf technology would, however, become particulary relevant if an international agreement were drawn up dividing the off-shelf zone between the interested countries. In that case, the "temporary measures" now announced would automatically lapse. But existing collaboration agreements between the Soviet Union and developing countries on mining activities normally provide for the Soviet Union's initial investment in the technology being repaid over several decades in appropriate minerals. **Vera Rich**

Conference all at sea

Washington

The largest ever diplomatic conference of the United Nations, with 160 governments represented, ended on 30 April in New York with a huge majority for a draft treaty on the Law of the Sea. But the participants in this eight-year-long conference have not gone home in triumph. Instead, they hope to limit the damage that has been done.

What went wrong? In the end, the United States voted against the treaty, as did Israel, Turkey and Venezuela. Britain, West Germany, the Netherlands, Belgium, Spain and the entire Eastern bloc abstained. Ironically, even the dissenters think the draft is a splendid basis for a treaty. It resolves several complex issues, such as navigation through narrow straits.

The stumbling block turned out to be the issue well advertised in the past few years: the rights of private companies to minerals from the deep sea bed. Both the United States and the Soviet Union were affronted. The new Administration in Washington had spent a year trying to decide how to amend the clauses on international access to oceanic minerals that the Carter Administration had negotiated. The Soviet Union, by contrast, objected to a form of words seeming to cast doubt on its freedom from free enterprise.

What happens now is anybody's guess. The procedure requires that governments wishing to sign the treaty must do so by December. Thereafter, the treaty will become international law when a specified quota of signatories has formally ratified it. The flaw is that because the draft of the treaty was not approved without dissent, it is open to any member of the conference to insist that the text cannot be a basis for international law because it was disputed.

So who shoulders the blame? The outcome of the conference was a failure for the diplomats. The bizarre handling of the United States negotiating team may have been crucial.

The Law of the Sea session just ended lasted just eight weeks, and brought to an end the year-long spell during which the United States Administraiton had put the negotiations on ice.

Private seabed mining has been the only stumbling block. Ideologically, the Administration objected to those parts of the text that would give developing countries a say over the private development of seabed minerals. But among the six objections listed by the United States when it returned to the negotiations in January were the provision that the treaty could be amended by three-quarters of its signatories without the US Senate's agreement, and the inclusion of the Palestine Liberation Organization among the members of the conference.

After the event, US negotiators claim that they were not given enough time to negotiate the real issues. The conference chairman, Tommy Koh of Singapore, made them stick to a prearranged timetable. The deadline of 29 March had early on been fixed for the filing of committee reports on which a final draft would be drawn up. By then, however, the mining committee had not agreed a resolution on seabed mining, while the constitutional objection to changing the treaty without consent of the US Congress had not been discussed.

Part of the difficulty seems to rest with the makeup of the US negotiating team. Its leader was James L. Malone, an assistant secretary at the State Department whom the Administration was stripping of his responsibility for nuclear non-proliferation. Malone's chief aide was Leigh Ratiner, a former government seabed mining expert working for the mining industry.

Ratiner was a free-wheeling negotiator and a veteran of earlier sessions at the law of the sea conference. Malone was by comparison a novice. Observers say that Malone would often say one thing and Ratiner another. The result would be either brilliant or confusing, depending on the observer. The delegation was under constant pressure from the mining lobby, which insisted that any sell-out to the conference would be denounced when the treaty came up for ratification in Congress.

The United States' demands for revisions of the treaty, circulated before the beginning of the last session in March, were quickly followed up by detailed textual proposals — and as quickly rejected as too sweeping. A compromise proposal, put forward by 11 states including Australia, Canada, Denmark, Finland, Iceland, Ireland, New Zealand, Norway, Sweden and Switzerland, was quickly rejected by the United States as inadequate but later recognized to be a basis for discussion. By then, however, it was too late.

What will happen now is far from clear. The ceremony at which the treaty will be opened for signature has been arranged for Caracas in December, but the host country, Venezuela, voted against the draft last month because if fails to resolve the dispute with neighbouring Colombia.

No state is bound by the vote it cast on 30 April. Abstaining states or no-voting states could still attend the signing ceremony. But Israel's no is unlikely to change, as the PLO no seems likely to stay in the final text. The Reagan Administration could not change the US vote without an ideological turnabout, which seems unlikely.

Ironically, the Administration's adamant defence of rights to ocean mining may have brought about a situation in which no mining is possible. Tommy Koh has threatened to sue any country that proceeds unilaterally to mine outside the framework of the Law of the Sea. So, if the United States, and other mining states such as the United Kingdom, remain out of the fold, they may see their mining ventures flee to other countries such as Japan.

Meanwhile Elliot L. Richardson, Law of the Sea negotiator for the Carter Administration, believes that the outcome of the conference is "sad". Richardson said that "ideologues" in and outside the Reagan Administration prevented the US negotiating team from being flexible.

The Administration, he said, "wasted" a year deciding whether to continue negotiating, and so spent too little time preparing for the talks themselves. At the meeting at which the US team agreed to produce specific draft articles, it had run out of time and had to include many items that it knew were unacceptable. Similarly, Richardson believes the US team should not have responded so negatively to efforts by a group of eleven countries to negotiate compromises. **Deborah Shapley**

0028-0836/82/190095-01$01.00

UK nuclear power

Bye British

At last the British nuclear power show is on the road. Advanced gas cooling is an also-ran in the UK Central Electricity Generating Board (CEGB)'s "statement of case" for the construction of a pressurized water reactor at Sizewell in Suffolk, published with due ceremony yesterday (Wednesday).

The price of the British 1,110 megawatt-electric power station is now estimated to be £1,147 million, excluding the initial charge of fuel, compared with £1,590 million for an AGR. The ratio is thus 72 per cent, substantially but not embarrassingly more than the 60 per cent Dr Walter Marshall, chairman of the PWR task force, was hoping for.

The task force was set up to control the cost of safety systems on the British PWR, which a year ago were escalating because of CEGB requirements to keep radiation exposure levels for PWR workers down to those expected of the cleaner gas-cooled reactor. Then there were fears that a British PWR might cost as much as an AGR. In the event, CEGB has maintained its safety requirements, adding elements — such as a larger containment building and a double skin — to the task force design; but the task force effort has shown up in the final price.

An equivalent coal-fired station would cost about the same as the PWR (£1,080 million), says CEGB, but fuel costs would be so much higher with coal that a PWR would give a net saving of some £500–£1,100 million over its 35-year life.

But these few lines give no indication of the scale and density of CEGB's argument which in the statement of case, the accompanying pre-construction safety report, the reference design and the supporting documents amounts to 2½ hundredweight (125 kg) of reading matter.

PWR objectors will now have eight months to study the case, before the public planning inquiry which is to be held at the Maltings, Snape, in Suffolk, in January 1983 — and they will need all of that time. Moreover, CEGB promises a supplement in two months' time which will give a sensitivity analysis (or estimate errors) on the economic elements of the report, and in late June or early July the Nuclear Installations Inspectorate is due to publish its reaction to CEGB's design. The opposition could thus find itself flattened by mere weight of words.

The report, and preparation for the inquiry (including fees to counsel), will cost CEGB some £5–10 million, it is estimated. Some copies will be made available to *bona fide* enquirers, 15 will be placed on public display and others will be sold at £250 a time.

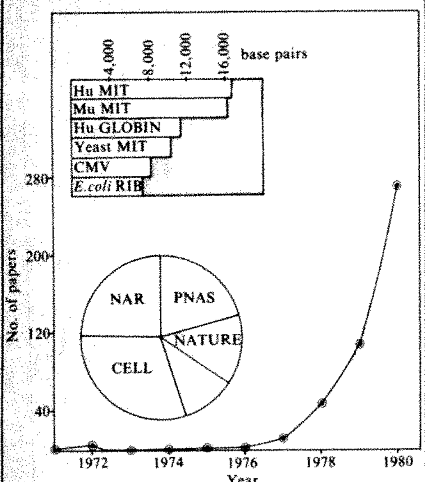
The next crucial task for CEGB is to define the board's procurement policy — where it will buy the components and design work for the reactor. The board estimates that 90 per cent of the PWR cost will pay for work done in Britain — but since the nuclear component (the nuclear steam supply system) amounts to only 10 per cent, CEGB is clearly not ruling out the possibility that the whole of the system might be bought abroad.

The reactor pressure vessel will certainly be bought abroad. Tenders are now out with Combustion Engineering of the United States and Framatome of France. For the rest of the system (such as the steam generators), discussions are beginning with potential British suppliers. Brian George, CEGB's PWR director, said it would be "too strong" to say that the board saw no reason why it should not buy British. In other words, there are good reasons — of price and confidence — why the board should buy foreign, and it will be up to British nuclear industry to convince CEGB otherwise.

Robert Walgate

Sequences add up

The first newsletter of the newly established nucleotide sequence data library of the European Molecular Biology Laboratory (see *Nature* 15 April, p.596) lists 565 entries and is claimed to be fairly complete up until 1980 and perhaps half complete for 1981. Four journals, *Cell, Nucleic Acids Research, Proceedings of the National Academy of Sciences* and *Nature,* dominate the entries. The data base contains nearly



600,000 nucleotides, with 18 continuous sequences of greater than 5,000 nucleotides. The longest of these, at 16,569 nucleotides, is that of the human mitochondrion (HuMIT). Mouse and yeast mitochondria and a human globin gene cluster also provide sequences of greater than 10,000 nucleotides and cauliflower mosaic virus (CMV) and *Escherichia coli* ribosomal genes are not far behind. The mouse provides the biggest number of entries, followed by man. In both cases the sequences are largely of immunoglobulin, globin and interferon genes. The newsletter is available from G. Hamm, EMBL, Postfach 10 22 09, D-6900, Heidelberg, FRG.

Alternative energy

Winding up

Waalre, The Netherlands

The Dutch are again turning to wind as a source of energy. At the end of the last century, more than 9,000 windmills were pumping water or grinding flour in the Netherlands, and if the present emphasis on wind energy continues, by the year 2000 there will be more than 2,000 megawatts of electricity produced from wind power. This is equivalent to about 13 per cent of the country's present requirement for electricity, which is not expected to increase much in the next few decades.

"Looks like the wind's dropped again"

Approximately two-thirds of the wind-derived electricity at the end of this century is expected to come from purpose-built wind-energy parks, with the rest from up to 15,000 smaller individual turbines. There are already 50 small turbines in use, manufactured by 7 Dutch and 14 foreign companies. Until now the growth of the use of wind energy has depended on the enthusiasm of local authorities, but the end of the national five-year research programme and the move into a development programme should see a further boost to wind energy.

The national research programme saw the construction of a 300-kilowatt test turbine, and the new programme will initially involve the building of a wind-energy "farm" with forty 250-kilowatt turbines as a prototype for much larger central electricity generating areas. The cost of this first "farm" is estimated at 40 million guilders (£8 million), half to be paid by the government and half by the public utilities companies.

The government is encouraging the construction of wind turbines by providing a subsidy of up to 40 per cent, which can be paid back over 10 years. The national development programme will also involve the design of a large 60–70 metre commercial wind turbine. Further research is to be done on tip-vanes and a multiple wind turbine is under development, comprising a tower with three cross-bars, each of which will support two rotors, giving a total capacity of 320 kilowatts.

Casper Schuuring

0028-0836/82/190096-01$01.00

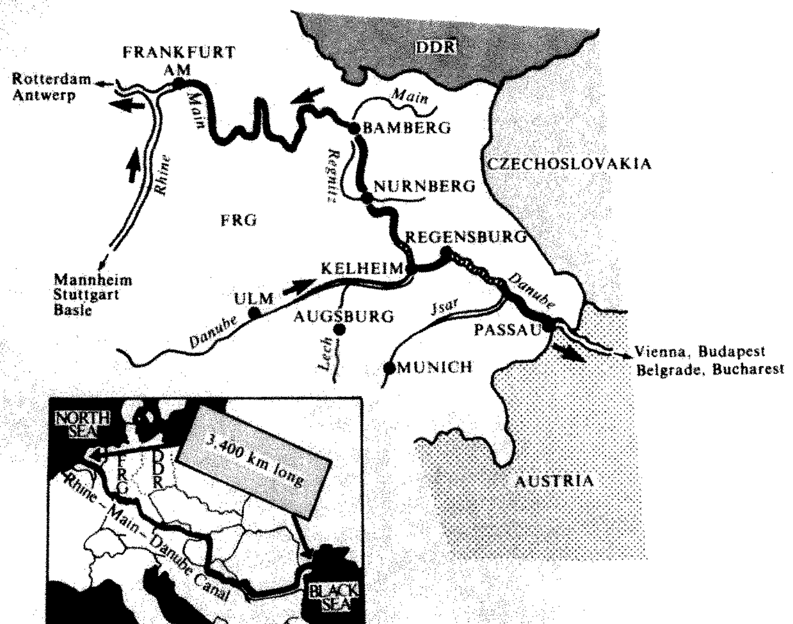
West German waterways

Canal in limbo

Heidelberg

Of the 3,500 km of waterway linking the North Sea to the Black Sea, only 55 km of the Rhine–Main–Danube Canal and 59 km of canalization on the Danube are still incomplete. However, as the work forges on, money is tighter, and controversy about the impact of the canal increases.

The first attempt to conquer the European watershed by joining the Rhine to the Danube dates back to Charlemagne. But it was not until 1855 that, under Bavarian King Ludwig I, a 15-m wide canal with 100 locks was opened. This never competed successfully with the railways and the last 120-ton barge passed through in 1945. Contracts from 1921 and from the 1960s commit the federal government to two-thirds and Bavaria to one-third of the costs of the present canal from Bamberg on the Main to Kelheim on the Danube.



Even by 1976, estimates showed that on each deutschemark invested, the profit would be only 50 pfennig. Total German water freight is in any case falling off: estimates for the canal were 14 million tonnes per year in 1969 and now stand at 3 million tonnes per year. Most of this would otherwise go by the *Bundesbahn*, which is predicted to lose DM60–120 million per year. Worse, German national carriers could face competition from state-subsidized East Europeans not only on the new waterway, but also in the crucial Rhine system. Furthermore, north German seaports could suffer when the canal links Austria, Hungary, Bulgaria and Romania directly to the Rhine delta.

Bavaria considers the project essential to combat local unemployment and underdevelopment and that the DM 3,150 million already invested should not be jeopardized. However, the DM 1,750 million for completing construction and the DM 12 million per year for running the canal might benefit the area more if spent differently. More jobs could be lost in farming and tourism than are created by the canal.

The plans also involve pumping water at 15 m^3 per second up from the Danube to flush out the severely polluted Regnitz and Main and provide cooling water for the four new power plants planned for the Main valley. Without the canal, a DM 500 million pipeline would be needed. But, integrated energy planning and improved waste treatment could meet these needs.

To the 55-m width of the canal will be added a huge parallel service area which in places will carpet the valleys of the Altmuhl and Sulz from wall to wall. After canalization the Danube will deepen its bed, lower the water table and drain surrounding wetlands. These vast landscape changes are the target of mounting environmentalist attack. The Altmuhl valley and the wetlands north of the Sulz and east of Regensburg are not only exceptionally beautiful, but also contain internationally significant biotopes. The destruction of the migrant wintering and resting grounds, the habitats of 46 endangered bird species and 13 endangered fish species, is contrary to the spirit of both Bavarian and federal laws framed in the 1970s to protect fast dwindling wildlife resources.

These serious environmental objections apart, the hard core of the current controversy is economic. In September 1981, while West German Chancellor Helmut Schmidt reassured Kreisky that work on the canal would continue, federal transport minister Hauff, on the strength of his latest reviews, was talking about "the dumbest project since the Tower of Babel". In January the cabinet instructed Hauff to ask the Bavarian Administration for talks about halting or reducing the scheme. Bavaria is holding firmly to its contracts, but Hauff and Strauss will meet this month. Meanwhile, there is nothing in the federal kitty for next year. **Sarah Tooze**

Kew Gardens

Changes ahead

In the cold climate of recession, the Royal Botanic Gardens at Kew in Surrey continue to flourish. Today (13 May) the Queen will re-open the Temperate House after a 5-year renovation programme costing £1.6 million. Building of this vast glasshouse, designed by Decimus Burton for the 3,000 different types of plant from the warm, temperate regions of the world, began in 1860.

The Royal Botanical Gardens may themselves soon undergo a historic restructuring. Since 1903, the institution has been part of the UK Ministry of Agriculture, Fisheries and Food (MAFF) but now talks are being held between officials from the ministry and Kew Gardens to transfer management to a body of trustees. Organizations with an interest in the gardens have been asked to submit their views by 21 May.

The change to trustee status would be largely managerial — the gardens would continue to be sponsored by MAFF. But this arrangement would permit a degree of adminstrative autonomy the lack of which has irritated past directors of Kew. The new board of trustees, many of whom would be scientists, would take decisions on the direction and development of policy. In April, Mr Peter Walker, Minister of Agriculture, Fisheries and Food, told the House of Commons that the present arrangements "are not ideally suited to the management of an institution which combines the functions of curation, research, advice and instruction and public amenity". At present the director, Professor Arthur Bell, is advised by a 12-strong Scientific Advisory Committee drawn mainly from university botany departments. Members of the Forestry Commission and the National Trust sit on a consultative panel at Wakehurst Place, the satellite garden in Sussex.

Kew Gardens have not experienced cuts as such but staff has been reduced through natural wastage from 480 in 1979 to the present level of 440 — a rate which compares with cutbacks at the ministry itself. Total running costs are £6 million, of which £3 million is accounted for by salaries.

Like London Zoo, Kew is a blend of scientific institution and tourist attraction. With an entrance fee of only 10 pence, and over 1 million visitors a year, however, the gardens could not survive without exchequer support. Some 3,100 genera of 340 families are represented in the living collections at Kew, and there are 5 million dried specimens in the herbarium. The physiology section at Wakehurst Place now incorporates a seed bank. But the aesthetic way of displaying plants of scientific and economic interest may obscure the extensive scientific work that the institution performs. **Jane Wynn**

0028-0836/82/190097-02$01.00

CORRESPONDENCE

The Johns Hopkins and India

SIR — In the news section of your 11 March issue, K. S. Jayaraman from New Delhi makes several serious and erroneous allegations regarding the Johns Hopkins University and its participation in research projects in India. I believe it to be important that the record be corrected.

For many years, scientists from the Johns Hopkins University collaborated with Indian colleges in two major projects. One was the pioneering research conducted from 1961 to 1974 at the Narangwal Rural Health Research Centre in the Punjab. The purpose was to design and evaluate practical approaches for the delivery of primary health care. A series of projects under the Indian Council of Medical Research involved a large group of Indian scientists and a few foreign scientists under the direction of Professor Carl Taylor. These studies served as a prototype for other similar health services research subsequently conducted in many countries throughout the world. Professor Taylor, a member of the Institute of Medicine of the National Academy of Sciences and a frequent consultant to the World Health Organization, the World Bank and many other national and international organizations, is universally recognized as an authority in the complex problems of primary health care delivery. He was born in India and has spent half of his life living and working in the villages of the subcontinent. Contrary to Mr Jayaraman's report, the Indian government has officially confirmed that he has never been expelled from India nor has he been forbidden to come to India or any other country in a professional capacity. His visits to India since the Narangwal Project closed have been in response to official invitations. The allegation that the purpose of the Narangwal study was to provide a base for spying on the Halwara airbase is wholly false, indeed preposterous.

The second project, based at several national institutes in Calcutta, was under the direction of the late Professor Frederik Bang. It was one of a group of International Centers for Medical Research and Training which were funded by the National Institutes of Health. With Indian scientific colleagues, a large number of important investigations were conducted dealing with infectious diseases and nutrition. Indeed, many of the fundamental studies which have led to the now universally used methods for treatment of cholera and other severe diarrhoeas with oral fluid therapy were made by the Calcutta group. This laboratory and its scientific staff had no connection whatsoever with the Fort Detrick biological warfare laboratory. There was also no link with the US Navy except for informal, collegial scientific exchanges with the Naval Medical Research Unit, then based in Taipei, which was similarly actively engaged in endeavouring to find more effective and practical methods for the treatment of cholera.

Since its founding in 1916, the Johns Hopkins School of Hygiene and Public Health has been deeply involved in education and research in international health problems in Baltimore and in many countries throughout the world. It continues today as one of the largest institutions so concerned with international health. Its numerous alumni hold eminent positions in national and international organizations around the world. The false aspersions cast on the the institution by Mr Jayaraman are deeply resented.

D. A. HENDERSON
(Dean)

The Johns Hopkins University, School of Hygiene and Public Health, Baltimore, Maryland, USA

Ban the billion

SIR — A.J. Southward (*Nature* **294**, 215; 1981) writes: ". . . estimated six billion tons of oil that reaches the sea each year". as he comes from Plymouth, Devon, we have a right to expect that "billion" means 10^{12}: but in context that is surely impossible. We are therefore left to wonder whether he writes in American from an English address. Unfortunately, even "6×10^9 tons" of oil per year flowing into the sea sounds fantastic, given that the world's production of crude oil is about 3×10^9 tons per year. Does that mean that most of the amount talked about is natural seepage? The rest of the article makes that sound implausible. We are left with the uncomfortable feeling that, as the author has introduced one uncertainty of a factor of 1,000, he may just have got the figure wrong.

The confusion is worse confounded by the difficulty of distinguishing spoken "m" and "b", especially on the telephone. As a result, "million" and "billion" are frequently confused, particularly in newspapers. The uncertainty is thus increased to 10^6.

Sir, could I prevail upon you simply to ban the word "billion" from *Nature*? It seems to me unnecessary; given the enormous uncertainty, it would be much better simply replaced by "10^9" or "thousand million".

EDWARD EISNER

Department of Applied Physics, University of Strathclyde, Glasgow, UK

A. J. SOUTHWARD REPLIES — *Mea culpe:* Professor Eisner is quite entitled to his little bit of fun. The word should have been million — 6 million tons of oil per annum into the sea. I agree that the word billion needs replacing.

PhD applications

SIR — The leading article in *Nature* of 15 April (p.592) implies that "promising people from less favoured universities" find it hard to enter PhD courses because the quota system of awards puts graduate student selection into the hands of the luckier university departments.

I expect that supervisors usually want to choose the most productive and biddable graduate students quite independently of their college of origin. If this is so, the main barrier to students moving to a new institution for their graduate studies must be inadequacy of information about projects and places available elsewhere.

For several years this department (which has always had a strict policy of not allowing its own honours students to continue as PhD students here), has organized a scheme for exchanging this information amongst about 80 UK life science departments — mainly in biochemistry, molecular biology, microbiology and cell biology. The scheme also promotes a common application form and loosely coordinated timetable for selection.

This year, for the first time, we have published the *Compendium of Research Topic Outlines*, which forms part of the scheme, as a single volume of over 400 pages, with name and subject indexes. The first indications are that this has proved very useful to students in finding out where they can carry out the kind of research of most interest to them. In the future, the availability of this compendium to libraries and career advisers should make the system of greater use to students in non-participating departments.

I shall be glad to answer enquiries from anyone who thinks that the compendium, or the scheme as a whole, could be useful to their students.

I think that many of the participating departments now find the scheme indispensable, as we do. For this reason I have never understood why physical chemical and social science departments do not cooperate in a similar way.

A. F. W. COULSON

Department of Molecular Biology, University of Edinburgh, Edinburgh, UK

Man mismeasured

SIR — Scientifically I am certainly no supporter of the main thesis of Stephen Jay Gould concerning punctuated equilibrium. Thus I might be expected to be in favour of (or at least to watch passively) virtually anything that helps to reduce Gould's immense influence and effectiveness, especially if prepared by an outsider to the evolutionary arguments. But in truth, I found the personal attack on Gould by your reviewer of *The Mismeasure of Man* (*Nature* 8 April, p.506) positively nasty, unprofessional, with but slight reference to substantive issues (and then often granting Gould's correctness), and just plain ugly. Who is Mr Blinkhorn to tell us that the book has "the routine flavour of Radio Moscow news broadcasts", as though in and of itself even if it were true (which is not in any way documented), it should damn the book for ever. My reaction is that if Blinkhorn is correct, more of us should join him in listening to those radio waves. Blinkhorn ends by admiring Gould's skill in presentation, and adding "but what a waste of talent". For his part, Blinkhorn has shown us a remarkable lack of skill, unfortunately presumably making full use of *his* talent.

THOMAS J.M. SCHOPF

Department of Geophysical Sciences, University of Chicago, Chicago, Illinois, USA

NEWS AND VIEWS

Evolution of bowerbirds' bowers: animal origins of the aesthetic sense

from Jared M. Diamond

IN September 1872 the explorer Odoardo Beccari became the first European to ascend the Vogelkop mountains of New Guinea and to observe there the Arfak tribesmen in their villages. Among the many sights that impressed Beccari were the beautifully decorated little huts scattered in the forest. Each hut was in the form of a cone up to eight feet long and four and a half feet high, so tightly woven of sticks as to be rain-proof. In front of each hut was a little garden, consisting of a mossy lawn decorated with over a hundred red, blue, brown, black, yellow and green objects carefully grouped by colour. The objects that Beccari saw were natural and collected from the forest: fruits, flowers, mushrooms and beetle skeletons. The displayed flowers were fresh and were removed and replaced as they faded. A century later, when Europeans had become regular visitors to the Vogelkop, the decorations included coloured man-made objects, such as shotgun cartridges and yellow cardboard Kodak film cartons.

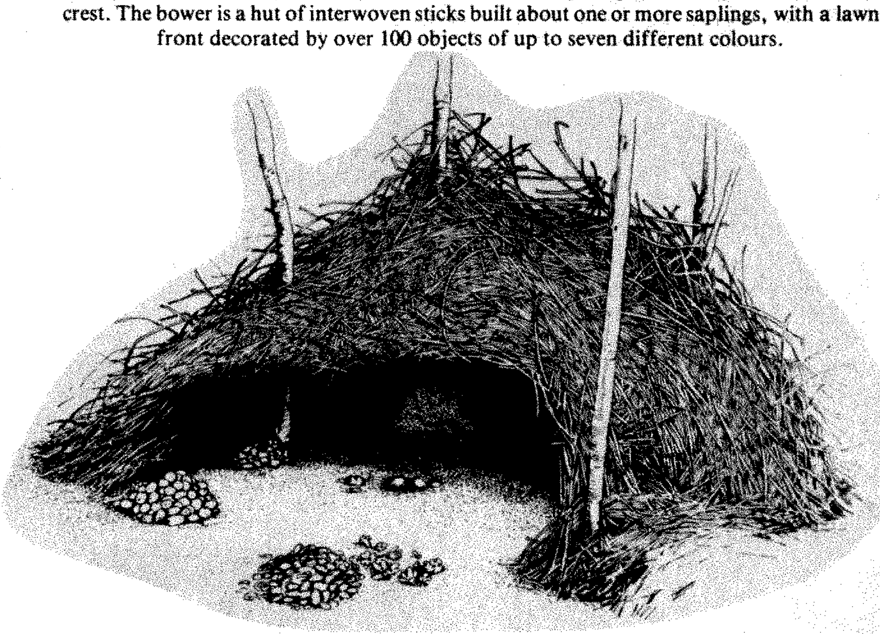
Was this a quaint native custom of the sort that draws anthropologists to New Guinea? Not at all. What caused a sensation when Beccari[1,2] published his account of these huts was that they were constructed, decorated and used by a bird, the Vogelkop Gardener Bowerbird.

The eighteen species of bowerbirds (family Ptilonorthynchidae) of which fourteen others besides the Vogelkop Gardener build bowers, are confined to New Guinea and Australia[3-5]. Bower structures vary with the species and include, besides the hut of the Vogelkop species, stick towers up to eight feet high, walled avenues, fern mats and moss platforms with parapets. Almost all bowers are decorated with coloured objects such as fruits, flowers, shells and bones, the colour preference varying with the species and with the individual bird. More diverse decorative objects appear in bowers near human habitation, and have included car keys, bottletops, tooth brushes, clothespins, false teeth and spectacles. At least six species use a tool to paint the bower yellow, green, blue, brown, or black with crushed fruit or charcoal, and at least four species orient the bower in a constant compass direction. Bowers are built by adult males and used as the site for wooing and inseminating females, which then fly off and do all the work of building a nest and rearing the young.

Understanding of the function and evolution of the bowers has come slowly, because the areas of New Guinea and Australia where many of the species live are remote and difficult of access. Indeed, the home grounds of a New Guinea species (*Amblyornis flavifrons*), known only from three skins sold to Lord Rothschild by a plume merchant in 1895, remained undiscovered until 1981 (see *Science* 23 April 1982), and the bowers of two other New Guinea populations (*Sericulus bakeri* and *Archboldia p. papuensis*) are still unknown. Two recent papers by Reta Vellenga[6,7] of Castle Hill, Australia, in conjuction with an earlier paper[8], provide the most detailed life history information yet available for a bowerbird.

Vellenga studied the Satin Bowerbird (*Ptilonorhynchus violaceus*) of south-east Australia, following on earlier work by Marshall[3], Chaffer[9] and Green[10]. The bird is about twelve inches long, the adult male glossy bluish-purple, the female and young male green. The bower is a woven platform about 10 feet square, supporting an avenue of woven sticks, with walls a foot high nearly joining in an arch over a floor, and decorated with blue or green natural or man-made objects. Perishable decorations such as flowers are replaced daily. Some bowers are left unpainted, but others are painted daily either blue, black, or green by the male, using a wad of bark as a paint brush and using crushed fruit, charcoal, or (nowadays near civilization) stolen blue laundry powder as paint. The long axis of the bower is generally within 30 degrees of north–south, possibly so that the male and female can face each other during early morning displays without either having to stare into the rising sun. When Marshall picked up a bower and reoriented it, its architect promptly demolished it and rebuilt it in the correct orientation. Between 1965 and 1971 Vellenga banded 940 bowerbirds in her backyard and followed the individual lives of many of them.

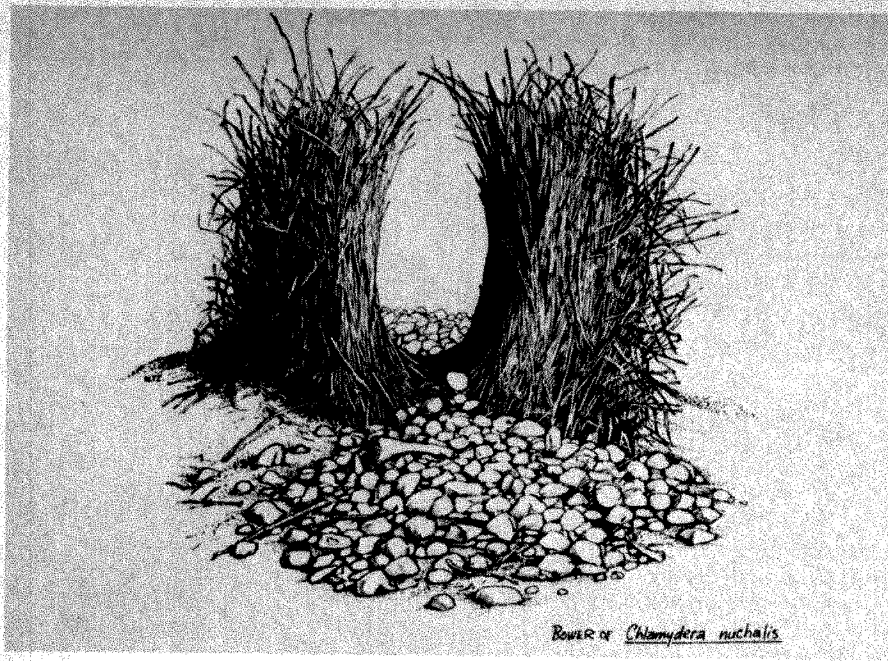
Extremely elaborate maypole bower built by a completely unornamented male bowerbird, *Amblyornis inornatus* (mountains of New Guinea's Vogelkop Peninsula). The male lacks a crest. The bower is a hut of interwoven sticks built about one or more saplings, with a lawn in front decorated by over 100 objects of up to seven different colours.



Drawings by W.T. Cooper (ref. 5).

Jared M. Diamond is Professor of Physiology at the University of California Medical School, Los Angeles, California 90024.

0028-0836/82/190099-04$01.00



Example of an avenue bower, built by the bowerbird *Chlamydera nuchalis* (Australia). An avenue is constructed of a floor and two painted walls of sticks and decorated with many stones and bones. Mating takes place in the avenue.

According to Darwin's theory of sexual selection, intraspecific competition for mates may select for ornaments or behaviours unique to one sex through either of two mechanisms: the value of the ornaments in attracting the opposite sex or in dominating rivals of the same sex. Bowers have traditionally been interpreted in the former sense, as love parlours where males seduce females. Vellenga's observations provide ample evidence for this function. The bower-owning male Satin Bowerbird continually tries to entice females into his bower by picking up in his bill an object such as a flower or snail shell, and posturing, dancing and displaying to the female. Rarely, the female is persuaded to enter the bower and copulation ensues. Far more frequently, the female flees without mating.

Why do courting males have such a low success rate? Perhaps females scrutinize various males and bowers before choosing. Many courtships are interrupted at a crucial stage by the intrusion of other individuals. However, another reason for the low success rate may be the invariant sequel to successful mating: the male bowerbird savagely attacks the female, pecks and claws her, and chases her from the bower. Mating itself is so violent that often the bower is partly wrecked, and the exhausted female can scarcely crawl away. The courtship display can appear little different from the male's aggressive display. When a courted female is won over and starts to solicit copulation, the male often changes his mind and chases her away. Thus, a female may have to make many visits to a bower before she overcomes her fear of the aggressive male. After mating, the female constructs a nest at least 200 yards from the bower and bears sole responsibility for feeding the young. Once the young leave the nest, the young of several females are tended by their mothers in a common nursery area. After the first year, bowerbirds outside the breeding season form separate flocks according to age and sex (females, young males and old males separate). Adult males that leave for the winter apparently dismantle their bowers first.

While Vellenga's studies confirm the traditional view of bowers as love parlours, they provide strong support for a simultaneous role of bowers in male–male interactions. Specifically, bowers, like flags of possession, may serve as symbols of males' property rights in their wars with other males. An adult male spends much time repairing his own bower, protecting it from raids by rivals, and attempting to steal ornaments or destroy rivals' bowers. bowers.

The battles and territorial shifts that Vellenga recorded among her 426 banded adult males make the European Thirty Years' War seem straighforward by comparison. When about 5 years old, male no. 85 won ownership of Vellenga's garden and maintained his bower there against intruders for 15 years. On 15 November 1967 he became ill, and other males demolished his bower, but by 28 November he had recovered and rebuilt the bower. Finally, on the morning of 26 October 1970 he became ill again and disappeared around 10 a.m. By 1:50 p.m. on that day his bower had been destroyed and two other males were in the vicinity. By 3 p.m. male no. 345, who had owned a bower 1.25 km to the west since at least April 1970, was in possession of Vellenga's garden and began constructing a new bower around 4 p.m. Male no. 345 retained Vellenga's garden through constant fighting until defeated on 3 January 1971 by male no. 460, who had until then owned a bower 7 km distant. Male no. 345 thereupon returned to his old bower, expelled a male who had taken it over and was still there when visited three years later, while male 460 still owns Vellenga's garden at latest report. In the meantime male no. 553 took over the bower of male no. 320,0.75 km south-west of male 345, while male no. 429 seized a bower 1 km to the south-east, and male no. 124. . .

What is the significance of all these details of bowerbird military history? As with other bird species, male bowerbirds vary in dominance, and the south-east Australian landscape varies in production of food resources consumed by bowerbirds. Superior locations such as Vellenga's garden are know to have supported active bowers for at least 30 years, though with a sequence of owners. Young males search for good sites to built a bower, and males that already own bowers spend some time searching for better sites. Thus, the old male 85 had long been dominant in Vellenga's neighborhood, and on his disappearance local dominance did not pass to a male at some different site but continued with the new owner of male 85's estate.

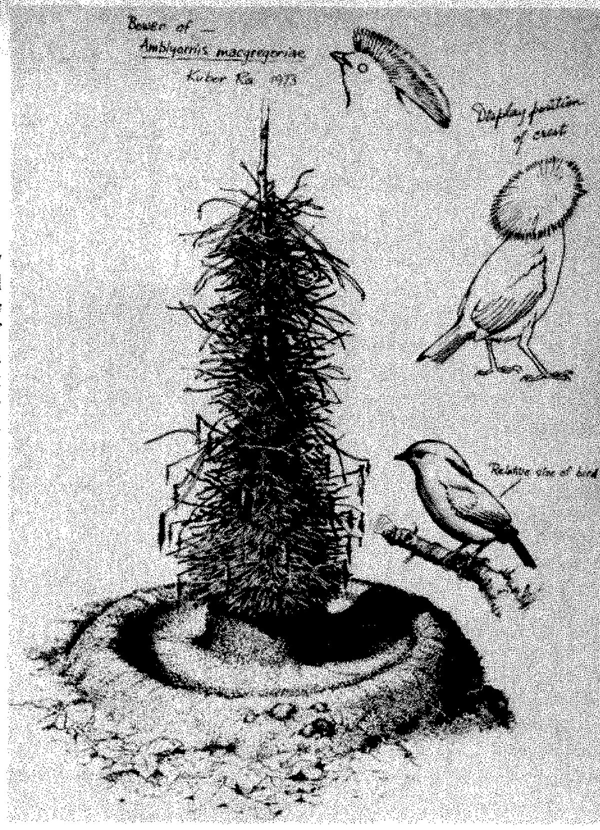
Dominant males directly prevent other males from wooing females by the destruction of their bowers. Young males continually try to erect rudimentary bowers in the territory of an adult male but the latter patrols his property several times a day and wrecks them. When Marshall placed 100 pieces of numbered blue glass into these rudimentary bowers one night, he found 76 of the pieces transferred to the bowers of dominant adult males by noon of the next day. Similarly, Vellenga observed a blue celluloid band to be transfered between bowers several times a day, until one male firmly wove it into his bower.

Dominant males may perhaps even affect their rival's chances through hormonal effects related to dominance and subordinance. The successful male bowerbird needs not only a bower but also blue plumage. Young male bowerbirds have green plumage resembling adult females and only begin to acquire the blue plumage of the adult male in their fifth year. In many animals the acquisition of adult sex-specific colour pattern is controlled by sex hormones, and domination of a male by rivals inhibits production of male sex hormones — it is possible that the acquisition of blue adult plumage by male Satin Bowerbirds is retarded in individuals that are continually harassed and see their bower-building attempts frustrated by dominant males.

How has the evolution of bowers taken place? Gilliard[4,13] argued that the female's attention during the male's courtship display has become 'transferred' from

male ornamental plumage to the bower and its decorations. The transfer is an advantage for the male bowerbird because it enables him to dispense with bright plumage that may attract predators. One line of evidence for this interpretation comes from comparison of courtship displays in the four species of the avenue-builder genus *Chlamydera*. In two species, *C. nuchalis* and *C. maculata*, the male has a lilac crest, which he erects and displays to the female by facing her while twisting his head to the side. In the crestless *C. cerviniventris*, the displaying male still twists his head as if to show the female a crest that his ancestors lost! But the male *C. cerviniventris* compensates by holding a spring of green berries in his bill during the display.

A second line of evidence for the 'transfer theory' is the inverse relationship, among bowerbird species, between elaborateness of male plumage and elaborateness of the bower and its ornaments. For example, among the five species of maypole builders, the male of *Amblyornis macgregoriae* has an erectile orange crest three inches long, builds a simple maypole four feet high and scarcely decorates it; male *A. flavifrons* has a gold crest of the same length, builds a similar maypole and decorates it only with three small piles of fruit of three different colours; male *A. subalaries* has an orange crest only 1.5 inches long but builds a dome over his maypole and decorates it with flowers, fruit, beetles and leaves of at least five different colours; male *Prionodura newtoniana* has a very short yellow crest but builds and decorates two maypoles up to eight feet high and joined by a bridge; and male *Amblyornis inornatus* is entirely crestless like the female but builds the elaborate bower described in the first paragraph of this article. A similar correlation exists among the eight species of avenue builders: from *Sericulus aureus* and *S. chrysocephalus*, of which adult males have golden and black plumage, decorate a two-walled bower only sparsely, and perhaps dispense entirely with a bower on occasion; to *Ptilonorhynchus violaceus* studied by Vellenga, with uniform blue male plumage and a painted two-walled bower with many ornaments; and culminating in *Chlamydera lauterbachi*, in which male and female are similarly dull but the male builds a decorated four-walled bower reinforced with up to 10 lb of stones.



Simple maypole bower built by an elaborately ornamented male bowerbird, *Amblyornis macgregoriae* (mountains of New Guinea). The male has a striking, erectile orange crest three inches long. A circular moss platform with a rim is constructed about the base of a sapling, about which sticks are criss-crossed to form a tower. The bower has few or no decorations.

The construction of bowers out of sticks and moss may have its origins in nest construction. Males of many bird species court females and induce them to ovulate by constructing true nests and offering food. It is possible that bowers began as courtship nests that became free of size constraints as the function of egg incubation was transferred to a separate nest built by the female alone. Decoration of bowers with coloured objects may have stemmed from courtship feeding, since displaying males of at least six species of avenue builder, one maypole builder, and the mat builder *Archbodia papuensis sanfordi* pick up a decorative object in the bill and extend it towards the female. In some cases this decorative object is an edible fruit, although female bowerbirds have never been reported to take and eat the fruit. If courtship feeding was indeed the original meaning of this gesture, the meaning has been lost in the many cases where the display fruit has been replaced by an inedible object, such as a stone, shell, or human artefact like a cloth or marble.

Eight other species build bowers which could less easily have had their origins in nest construction and courtship feeding. The bowerbird *Scenopoeetes dentirostris* rakes an area of ground clean and decorates it with dozens of leaves carefully turned upside-down; the four birds of paradise (family Paradisaeidae, not Ptilonorhynchidae) of the genus *Parotia* rake an area of ground clean without decorating it; and the two birds of paradise of genus *Diphyllodes* rake an area of ground clean and strip the leaves off the trees over this area.

Does bower building imply that bowerbirds are intelligent and have an aesthetic sense? It certainly seems true, as Vellerga concludes, that many skills related to bower building and use have to be learned. Young males in green plumage spend about two years building rudimentary but increasingly complex bowers before acquiring blue adult plumage and building complete bowers. These 'practice bowers' are the joint efforts of several young males, which take turns placing and rearranging sticks, often clumsily and without success, occasionally with the cooperation of more skilled older males. The young males do not paint these bowers and are less discriminating than adult males in choice of colour for bower decorations. The young males spend much time watching the displays, mating and other bower activities of adult males, and are displayed to by adult males.

As for aesthetic sense, colour preference in decorative objects varies with species and with individual, sometimes in a way that can be related to the bird's own colour. Adult males of *Ptilonorhynchus violaceus* prefer blue fruits, flowers, parrot feathers and man-made objects (see photograph on p.867 of ref. 10, depicting nine man-made objects recovered from bowers, all of them blue: a tooth-brush, toy airplane, pen, pencil-sharpener, clothes-pin, marble, shotgun cartridge, button and bottle-top). Blue is also the eye colour in both sexes, and is the colour of the adult male's plumage and bill. *Chlamydera caculata* lives in dry country, has brown plumage and eyes, selects pale ornaments and paints the bower brown. *Amblyornis flavifrons* displays with a blue fruit that affords maximum contrast with its golden crest, and consistently decorates the bower with fruits of three colours: the same blue fruit, green fruit that is the unripe form of the blue fruit, and a yellow fig similar to the colour of the abdomen and male crest. *Amblyornis inornatus* males vary individually: one used ornaments of seven different colours, another preferred brown and black, another favoured red and one specialized in white. The function of bower painting is mysterious, but a possible clue is that Vellenga twice observed a young male visiting a bower in the absence of its owner

and painting the walls with saliva, whereupon the owner returned, expelled the intruder and repainted the walls with green liverwort. Perhaps painting not only has an 'aesthetic' purpose but also provides an olfactory and visual mark of individual ownership, like the marking by urine or anal scent glands in many mammals.

Even if antecedents can be suggested for bower construction, decoration and painting, why have these display activities evolved to such an extreme degree in bowerbirds but in no other birds? One prerequisite is ample leisure time. Male bowerbirds spend most of their time repairing the bower, fighting rivals, raiding rivals' bowers and wooing females. Bowerbirds are among the largest frugivorous songbirds of Australia and New Guinea and are aggressive and dominant over other species because of their size. They may thus have preferential access to trees with concentrated fruit crops of high nutritional quality, enabling them to satisfy food needs in a short time and to devote remaining time to bower activities (see ref. 14). Other large tropical frugivores devote their leisure time to other bizarre social activities, such as the arboreal displays of most birds of paradise and the 'bowers' of the bird-of-paradise genera *Diphyllodes* and *Parotia* and the cotingid *Rupicola rupicola* (Cock-of-the-Rock).

It is apparent that major facets of bowerbird behaviour await further study. Gilliard argued that bowerbirds are lek species (ones with concentrated display grounds); this appears true for *Archboldia papuensis sanfordi* (known population confined to small areas on two New Guinea mountains), untrue for *Amblyornis flavifrons* (bowers spaced one-quarter of a mile apart) and unclear for other species. It is unknown whether simple, erratic, or nonexistent bower construction in the genera *Sericulus, Ailuroedus* and *Scenopoeetes* is a primitive condition or an advanced one: are these genera acquiring or losing the bower-building habit? Have bowerbirds co-evolved with certain tree species on whose fruit they specialize, as seems true for birds of paradise? Vellenga's work only begins to give tantalizing cues to the origin of colour preference, the origin of painting and the relative roles of instinct and learning in bower construction. To many ornithologists, bowerbirds remain the most intriguingly human of birds. □

1. Beccari *Ann. Mus. Civ. Genova* Ser. 1, 9, 382 (1877).
2. Beccari *Gardener's Chronicle*, 16 March, 332 (1978).
3. Marshall *Bower Birds: Their Displays and Breeding Cycles* (Clarendon, Oxford, 1954).
4. Gilliard *Birds of Paradise and Bower Birds* (Natural History Press, Garden City, New York, 1969).
5. Cooper & Forshaw *The Birds of Paradise and Bower Birds* (Collins, Sydney, 1977).
6. Vellenga *Emu* **80**, 49 (1980).
7. Vellenga *Emu* **80**, 87 (1980).
8. Vellenga *Aust. Bird Bander* **8**, 3 (1970).
9. Chaffer *Aust. Zool.* **12**, 295 (1959).
10. Green *Natn geog. Mag.* **152**, 865 (1977).
11. LeCroy *Am. Mus. Novit.* 2714 (1981).
12. Diamond *Nature* **293**, 257 (1981).
13. Gilliard *Scient. Am.* **209**, 38 (1963).
14. Snow *The Web of Adaptation* (Collins, London, 1976).

The missing mass — now it's a gravitino!

from Joseph Silk

COSMOLOGISTS have a dire secret. They have not the slightest clue to the nature of the dominant mass constituent of the Universe — only the inherent implausibility of the idea prevents them from postulating the existence of innumerable copies of *Nature* floating in space. Preference has alternated for many years between two candidates: black holes and degenerate black dwarfs of the substellar variety. The mass density in the invisible form must exceed that in luminous material by one to two orders of magnitude.

The amounts of matter and luminosity in a given region are commonly measured in units relative to the mass and luminosity of the Sun. One finds a mass–luminosity ratio of about 2 in the solar vicinity; this means that the matter around us consists of stars typically of somewhat lower mass (luminosity being very sensitive to mass) than the Sun. In a great cluster of galaxies such as Coma, the ratio is about 300 (if one specifies that blue light is measured and that the Hubble constant is 50 km s^{-1} Mpc^{-1}). Part of this difference can be understood in terms of the type of stars present in the cluster. Most of the cluster galaxies are ellipticals, which lack the young, hot, blue stars that contribute little to the mass of a galaxy such as ours but dominate its light. Thus, the mass–luminosity ratio of the underlying old disk and bulge populations of stars near the Sun that are most comparable with stars in ellipticals is about 6. Indeed, this is confirmed by measurements of the velocities of luminous K giants at a height of 1 kpc or more above the disk, which directly sample the local gravitational field. It follows that the total stellar content of the Coma cluster of galaxies amounts to only about 6/300 or 2 per cent of its dynamical mass.

Further, a substantial fraction (about 10 per cent) of the mass of the Coma Cluster has been discovered to be in the form of hot, X-ray-emitting gas. Thus, the total content of the Coma cluster that is in luminous form is only 12 per cent of its total mass. The remaining 88 per cent of the mass of Coma is invisible, at least hitherto, and its nature is a continuing source of speculation. A similar fraction of non-luminous matter appears to be present in the spiral-dominated groups of galaxies (although data on the intergalactic gas content of groups are very incomplete). These measurements both refer to scales of one or two million light years.

On smaller scales, galaxy rotation curves sample the distribution of dark matter in the outer parts of the galaxies. The predominance of flat rotation curves out to several times the optical radius of the galaxy indicates that the mass-to-light ratio is rising with increasing scale. The group and cluster data suggest that the invisible mass fraction of the Universe continues to rise at least up to scales of millions of light years. Whether there is still more dark matter unclustered on larger scales is unknown. One can say that for the Universe to be of critical density, corresponding to closure, the required amount of dark matter has a mass–blue luminosity ratio of about 1,000. Measurements of the redshift–magnitude (or velocity–distance) relation for distant galaxies suggest that the mean density of the cosmos cannot much exceed this critical value. This ratio is only a factor of 3 larger than the Coma cluster value.

Does this signify that the Universe is near its closure density? Almost certainly not, since rich clusters are rather rare objects. Far more typical is the sparse group which has a characteristic mass–blue luminosity ratio of about 50. The large values of mass to luminosity in clusters are due in part to the older (and dimmer) star populations, and in part to the fact that a substantial amount of matter that would perhaps have formed spiral disks is in the intergalactic medium. The phenomenon that is responsible for the dark matter is a universal one. Even studies of galaxy haloes where little light is seen indicate that the ratio of mass to luminosity must locally be extremely large, amounting to several hundred or more. The evidence that exceedingly dark matter dominates the mean mass density of the Universe seems overwhelming. Such matter exists in galaxy clusters and groups, in the haloes of galaxies, and even in the solar vincinity, where, however, the principal contributor to the density is the old star population.

Particle physicists have joined the quest for dark matter in the cosmos. Experimental indications (still unconfirmed) of a finite rest mass for the neutrino have led to a new candidate for the dark matter. The big bang theory predicts that the number density of neutrinos at present should be comparable with the number of photons, both being relics of the primordial fireball that described the early evolution of the Universe. If the neutrino has a rest mass of 1 eV (or one-millionth of an electron mass), it would presently dominate the mass density of the Universe. With a neutrino rest mass as large as 100 eV, the Universe could even be of closure density. The existence of a massive neutrino has notable implications for the large-scale structure of the Universe. Neutrinos are collisionless particles and cannot sustain density fluctuations. As a consequence, all small-scale structure in the Universe would be suppressed, until

Joseph Silk is Professor of Astronomy at the University of California, Berkeley, California 94720.

0028-0836/82/190102-02$01.00

very late times. This supports a theory of galaxy formation in which the galaxies arise from the fragmentation of large structures that have collapsed anisotropically into pancake-like sheets. Such anisotropies arise naturally in collapse when small-scale structure is suppressed, and lead to the formation of great voids between the clusters and superclusters of galaxies that develop in the high-density regions. One question that remains unanswered in such a theory concerns how the scale of a typical luminous galaxy is determined.

Now a new particle, the gravitino, has emerged from the creative minds of two particle theorists as a new candidate for the dark matter (H. Pagels and J. Primack *Phys. Rev. Lett.* **48**, 223;1982). The gravitino is a child of supersymmetry, associated with the search for a theory that unifies gravity along with the strong, weak and electromagnetic interactions. Supersymmetry unites fermions and bosons on a similar basis, and may help explain relations between certain large mass ratios involving the Planck mass, the mass associated with the grand unified theory symmetry breaking, and the mass associated with the electromagnetic-weak force symmetry breaking. Super-symmetry breaks down below energies of about 100 GeV, and results in the production of long-lived fermion partners of the graviton and the photon, the mediators of the only long-range forces. The new particles are called gravitinos and photinos, and both may have finite rest masses. The number of photinos is comparable with the number of neutrinos, and the implications of a finite photino mass are indistinguishable from those of a neutrino rest mass.

However, the gravitino uncouples sufficiently early that the predicted number of gravitinos is only about 10 per cent of the cosmic neutrino flux. Subsequent decouplings of other particles produce the bulk of the neutrinos. This leads to an interesting consequence: if gravitinos are of sufficient mass to account for the dark matter in the Universe, the gravitino rest mass must be some 10 times greater than the neutrino rest mass, or as large as 1,000 eV if gravitinos are at closure density. As the Universe expands, the more massive gravitinos become non-relativistic at an earlier epoch than do the neutrinos. The horizon scale at this instant provides a measure of the minimum primordial fluctuation scale that can survive the free streaming of particles at the speed of light. For neutrinos of mass 30 eV, the perturbation amplitude peaks at 4×10^{15} $M_{\odot}$, whereas for 1 keV gravitinos the perturbation spectrum peaks at about 10^{12} $M_{\odot}$. This is just the mass-scale associated with the dark haloes of galaxies. If we take the coincidence seriously, then gravitinos of mass 1 keV are the dominant source of matter in the Universe, and may account for the characteristic scale of galaxies.

Indeed, it turns out, according to unpublished calculations by J. R. Bond, A. S. Szalay and M. S. Turner, that gravitinos may lead to the best of all possible scenarios for large-scale evolution. For their mass spectrum extends all the way from scales of supercluster mass down to scales comparable to that of dwarf galaxies. Unlike the neutrinos, gravitinos do not dominate the mass density of the Universe until long after they have become non-relativistic, and this enhances the growth of large-scale gravitino perturbations relative to smaller-scale perturbations, whose growth is inhibited by the inertia of the radiation-dominated mass content of the Universe.

Dark matter may therefore consist of massive neutrinos or gravitinos, or perhaps some more exotic species of elementary particle whose presence arises from unification theories that describe the first 10^{-36} second of the Universe. Such cosminos are the bane of the astronomer's life, for they are likely to be forever invisible, yet they provide a potential solution to some of his greatest problems.

The regulation of contraction in cardiac muscle

from D.A. Eisner and D.G. Allen

One hundred years ago Sydney Ringer, a physician working at University College London, discovered that calcium is required for the heart to contract. Only in the past twenty years, however, have the intracellular mechanisms responsible for this calcium dependence become clear and the central role of intracellular calcium in the control of cellular processes become widely appreciated. In the past five years our understanding of the regulation of cardiac contraction has taken a leap forward with the introduction of techniques for measuring the crucial intracellular ion activities (Ca^{2+}, Na^+, H^+, K^+ and Cl^-) and the development of biochemical methods for assessing the regulation of protein function. The field of cardiac electrophysiology also seems poised for a period of rapid growth with the recent introduction of single cell and patch clamping techniques. Each of these fields was represented at a recent meeting* at University College London which concentrated on two aspects of the regulation of cardiac contraction. The first was that of how the low resting level of calcium is maintained and how the increase of intracellular calcium required for contraction is produced. The second concerned mechanisms other than changes of calcium concentration which may be involved in the regulation of cardiac contraction.

The role of Na^+/Ca^{2+} exchange in the regulation of intracellular calcium dominated the first half of the meeting. L.J. Mullins (University of Baltimore) suggested that the intracellular calcium concentration is controlled by a voltage-dependent Na^+/Ca^{2+} exchange which, at rest, pumps calcium ions out of the cell in exchange for sodium ions that enter by flowing down their electrochemical gradient. A novel aspect of his model is that it attributes much of the calcium entry during the long cardiac action potential to this mechanism. It was suggested, by analogy with experiments on squid axons, that depolarization makes the Na^+/Ca^{2+} exchange bring Ca^{2+} into the cell and extrude Na^+. In contrast, H. Reuter (University of Berne) stressed the importance of conventional Ca^{2+} channels in providing the influx of calcium ions during the action potential. Recent recordings from single channels have removed lingering doubts about the existence of the calcium channel in the heart. It is clear that quantitative measurements are now required to establish the relative contributions of Ca^{2+} channels and Na^+/Ca^{2+} exchange to the rise of Ca^{2+} which produces contraction.

Papers given by D. Eisner (University College London) and by D. Ellis (University of Edinburgh) were concerned with altering the magnitude of Na^+/Ca^{2+} exchange and examining the effects of contraction. It is clear that small changes of internal sodium concentration have enormous effects on contraction. These effects depend on the membrane potential and the external sodium concentration in a way which is, at least, consistent with the predictions of a voltage-dependent Na^+/Ca^{2+} exchange model. However, it was also shown that changes of internal and external sodium concentrations have pronounced effects on intracellular *p*H and these may also affect contraction.

Evidence that the contractility of cardiac muscle can be regulated by mechanisms other than changes in the calcium levels was also presented. Direct evidence was presented by D. Allen (University College London) who, using the bioluminescent protein aequorin as a calcium indicator, showed that as well as increasing the cytoplasmic calcium concentration during contraction, adrenaline reduces the sensitivity

*The Physiological Society Symposium on 'The Regulation of Contraction in Cardiac Muscle' was held on 25-26 March 1982.

D.A. Eisner and D.G. Allen are in the Department of Physiology, University College London, Gower Street, London WC1E 6BT.

0028-0836/82/190103-02$01.00

of the contractile proteins to calcium. In an examination of the biochemical basis of this phenomenon, P. England (University of Bristol) showed that the application of adrenaline results in a protein kinase-dependent phosphorylation of troponin and other proteins. The phosphorylation allows the troponin to speed up its release of bound calcium and may help heart muscle to relax more quickly — of obvious importance for the high heart beat rates produced by adrenaline. Other proteins which become phosphorylated in the presence of adrenaline include a component of the sarcoplasmic reticulum calcium pump and a surface membrane protein. These phosphorylations may therefore affect movements of calcium between the cytoplasm and the sarcoplasmic reticulum and between the cytoplasm and the extracellular space. It was suggested by F. Flitney (University of St Andrews) that some of these phosphorylations could be controlled by the relative levels of cyclic AMP and cyclic GMP.

A major difficulty inherent in studies of the regulation of contraction is that the intracellular calcium concentration cannot be controlled when the studies are performed on intact cells. Unfortunately, biochemical studies suffer from the disadvantage that cell constituents involved in the regulation of contraction may be lost. To circumvent this problem, S. Winegrad (University of Philadelphia) has attempted to make cells which are 'hyperpermeable' to calcium ions but which retain larger substances within the cell. Using this technique he produced evidence for the control of the calcium sensitivity of contractile proteins by phosphorylation of troponin. This approach came under attack from D. Miller (University of Glasgow) who suggested that such cells are not really permeable to calcium ions. It appears that this problem must be resolved before the results can be unequivocally interpreted. Winegrad also described recent experiments which suggest that a proportion of the myosin molecules are normally inactivated but can be activated by a cyclic AMP-dependent process. If confirmed, this raises the prospect of an additional mechanism whereby force production in cardiac muscle can be varied. □

Recent developments in optical storage technology

from Alan E. Bell

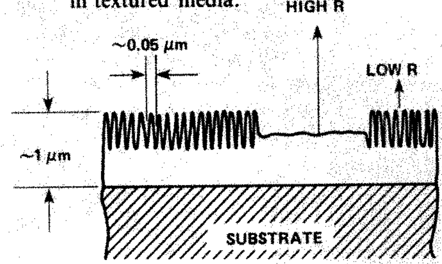
RESEARCH and development in the field of high-density optical data storage media and devices continue to move forward at a brisk pace — in the eighteen months since last reviewed in these columns[1], several companies have committed themselves to development efforts and the introduction of commercial optical disk computer memories can be expected in the next few years. Such devices will offer rapid random access (∼200 ms) to a database of up to 10 gigabytes per disk at a cost of 10^{-6} cents per byte or less and seem likely to leap-frog the projected improvements in the currently ubiquitous magnetic data storage media. The impact in many areas of electronic data processing, ranging from the 'electronic office' to large data-base archives and even to small business and personal computer systems, will be profound[2].

The two most critical components of the optical memory system are the diode laser read/write light source and the optical storage medium itself. The diode laser source is much preferred to the alternative gas laser sources, mainly because of the small size and power requirement, the capability for direct modulation of the optical record beam by the input data signal and the potentially high reliability and lifetime. Recent advances in the design and fabrication[3-8] of high-power (>40 mW output at 10 MHz modulation) diode lasers have made them available as commercial products, albeit at an initially high cost. With manufacturing scale-up and improved process yields, the cost should rapidly decrease to less than $100 (ref.9).

The 'ideal' optical storage medium has yet to be found and the past year has seen reports of many new recording media and configurations. Several new write-once (that is, irreversible) recording mechanisms have been identified and significant gains in the optimization of reusable (that is, reversible) media have been made.

Workers at Bell Telephone Laboratories (BTL) have reported the recording characteristics of irreversible media, microscopically textured low-reflectivity films of vacuum-deposited semiconductors[10]. The textured surface layer (see Fig. 1) is formed

Fig. 1 Recording mechanisms in textured media.



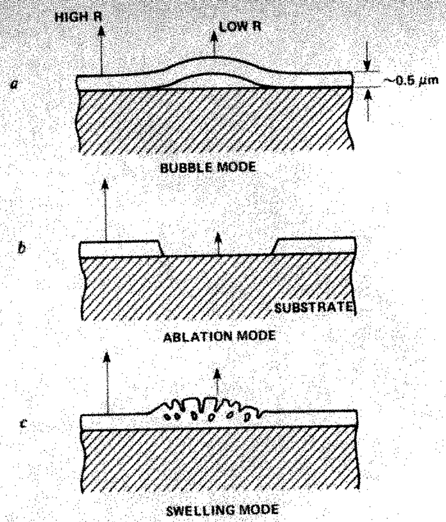


Fig. 2 Recording mechanisms in hydrogenated amorphous semiconductor films. *a*, Bubble formation; *b*, ablation; *c*, sponge formation

by plasma etching the semiconductor (Si or Ge) layer in the presence of aluminium. Atoms are sputtered from the aluminium surface and deposited on the semiconductor film to form a relatively thin and discontinuous layer of aluminium islands with dimensions of 100 nm or less. These act as an etching 'mask' that produces a microscopic needle-like texture on the etched semiconductor surface. The texture has a relatively broad-band antireflection characteristic (previously used in solar thermal collectors[11]) and can be locally melted by an efficiently coupled recording beam. Surface tension promptly restores a more nearly planar geometry to the melted area (see Fig. 1) and the data so recorded may be retrieved by detecting the increased reflectance of the smoothed region. Some critical problems of this approach do remain — it is not yet known how much playback signal-to-noise ratio (SNR) is reduced by the inherent 'graininess' of the textured surface, and whether the plasma etching can be sufficiently well controlled to act uniformly over disk-sized areas (∼0.1 m^2.).

Bösch[12], also of BTL, has investigated the recording characteristics of amorphous hydrogenated semiconductors prepared by reactive sputter deposition of the semiconductor material in a gas mixture containing 25 per cent hydrogen in argon at 10 mtorr. Three distinct recording mechanisms are possible, all of which are activated by the effusion of the hydrogen from the region heated by the laser beam; a process which occurs at a relatively low temperature and is strongly dependent on the choice of semiconductor. At low powers a blister or bubble region (Fig. 2*a*) forms in amorphous hydrogenated silicon semiconductors. At higher powers, ablation of the films occurs, apparently without melting, since no re-solidified residue is observed at the rim of the pit (Fig.2 *b*). In the case of amorphous hydrogenated germanium semicon-

Alan E. Bell is Senior Staff Physicist at Exxon Research and Engineering Company, PO Box 45, Linden, New Jersey 07036.

0028-0836/82/190104-03$01.00

ductors, the effusion of hydrogen produces a micrometre-sized porous region, somewhat resembling a sponge (Fig.2c). The hydrogenated semiconductors provide a very stable storage medium but recording with a diode laser source is difficult because the wavelength of the beam (~820 nm) is in the region of the bandgap energy and absorption is therefore low.

Bubble formation is perhaps the most intriguing recording mechanism reported by Bösch and has previously been described by workers at 3M[13] and Thomson-CSF[14]. In the Thomson-CSF approach, micrometre-sized blisters are formed in a thin layer of a noble metal alloy (for example, Au or Pt alloy) by partial vaporization of an underlying polymer layer. The scattering of the read beam by the blister results in a relatively low effective reflectivity of the blister compared with the undisturbed planar regions (Fig. 3*a*). Workers at 3M and others[15] have combined the bubble mode storage medium with the anti-reflection trilayer structure[1,16] (Fig. 3 *b*). The trilayer increases optical efficiency ($A \leq 0.9$) and thus improves sensitivity. When the blister forms in the uppermost (absorber) layer the interference properties which cause the structure to have low reflectivity are unbalanced and a high-reflectivity blister region forms. One advantage of the blister approach is its potential for extremely high SNR. Irregularly shaped residue at the pit rim, often the limiting factor in the playback SNR of melted pit media, are absent. Furthermore, since the recording sensitivity of bubble mode media is dictated by the thermal vaporization properties of the underlying polymer layer, rather than those of the bubble-forming absorber layer, relatively refractory and stable materials can be used for the latter without sacrificing the recording sensitivity normally associated only with low melting point, often unstable, recording materials[17].

A novel approach to optical storage media has been disclosed by workers at IBM[18]. A layer of amorphous silicon is coated with a noble metal layer, such as Rh or Pt, to a thickness which has a low reflectivity at the recording wavelength. Heat generated by the absorption of the recording beam energy causes the two layers to react and form a microscopic mark consisting of relatively highly reflective noble metal silicide (Fig.4). This approach provides archival stability but appears to be relatively insensitive. Unfortunately, if an attempt to improve sensitivity is made by choosing metals (such as Pd) which form silicides at relatively low temperatures, then the rate of reaction between the layers, even under room temperature storage conditions, leads to an unacceptibly low archival lifetime for the disk.

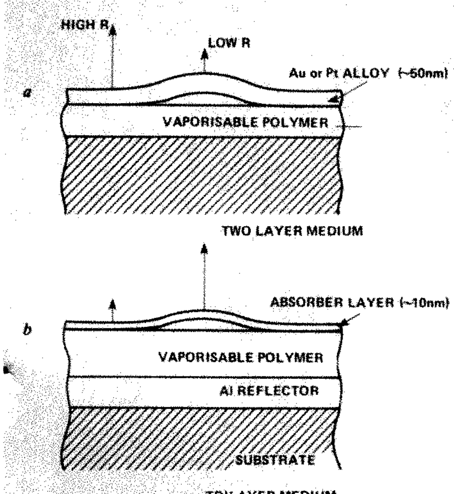


Fig. 3 Bubble mode recording media. *a*, Bubble formation in single layers; *b*, bubble formation in trilayer structures.

For a number of data storage applications, for example the filing of legal, financial and medical records, the write-once storage media discussed above provide a positive guarantee that the records have remained unchanged since input. Nevertheless, computer-operating systems have evolved for a period of 30 years with data storage devices that offer erasure and rewrite capability. Two potential candidates for reversible optical storage media that have been intensively investigated are the chalcogenide alloys such as TeGeAs[19] and magneto-optic alloys such as MnBi[20] and GdCo[21].

The chalcogenide alloys exist at room temperature in either the crystalline or amorphous form, each having different optical properties. An amorphous mark is recorded in a polycrystalline film by briefly melting (~50 ns) the crystalline material, and the subsequent rapid cooling (~10^{10}° s^{-1}) 'freezes' the random atomic structure of the melt to produce an amorphous solid. Recrystallization (that is, erasure) can be accomplished by laser beam heating the amorphous region to a temperature above the glass transition temperature but below the melting point. If the temperature is maintained for longer than about 1 μs, the amorphous dot is annealed and so returned to the original polycrystalline state. Bell and Spong[22] demonstrated reversible optical recording of video signals with an encapsulated tellurium trilayer disk. The encapsulation inhibits the irreversible process of hole formation during the molten phase of the amorphization step. Despite earlier reports[23] indicating limited cyclability of chalcogenide alloys it was found that with pure tellurium up to fifty record and erase cycles were excuted on a single track without significant deterioration of signal quality.

In contrast to the chalcogenide alloy phase-change materials, the number of cycles of magnetization reversal possible in conventional magnetic media appears to be unlimited. In reversible magneto-optic materials, magnetization reversal is accomplished by Curie-point writing or other techniques[24]. In Curie-point writing the magnetization of a micrometre-sized region of the film is reversed by first using the record

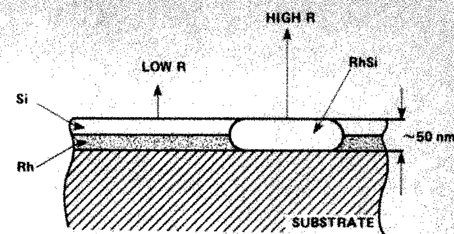


Fig. 4 Alloying bilayer approach.

laser beam to heat temporarily the region above the Curie temperature and then cooling in the presence of an external magnetic field oriented antiparallel to the initial (perpendicular to the substrate) magnetization direction. The presence of the magnetization reversal is detected by the sign of the Kerr or Faraday polarization rotation produced in light respectively reflected from or transmitted by a magnetized layer. In the past, playback SNR has been limited by the low optical signal contrast between oppositely magnetized regions. It has been known for some time[25], however, that the Kerr effect, in particular, can be increased by coating the magneto-optic material within an anti-reflection structure[16]. Mansuripur *et al.*[26] showed that the incorporation of a magneto-optic material into the trilayer structure can lead to a significant increase in signal contrast and thus SNR on playback. The incorporation of recently developed ternary amorphous magneto-optic alloys such as TbGdFe[27] into trilayer structures promises the practical realization of reversible high-density optical data storage. Incidentally, the interference properties of the trilayer structure can also yield significant increases in SNR of both Raman and Kerr spectroscopy used in basic studies[28,29] of the properties of very thin films ($\leq$ 10 nm).

1. Bell, A.E. *Nature* **287**, 583 (1980).
2. Goldstein, C.M. *Science* **215**, 862 (1982).
3. Ueno, M. *IEEE J. Quantum Electron.* **QE-17**, 2113 (1981).
4. Ettenberg, M. & Botez, D. *Electron. Lett.* **18**, 153 (1982).
5. Botez, D., Channin, D.J. & Ettenberg, M. *SPIE Proc.* **321** (in the press).
6. Kawabe, H. *et al. 13th Jap. Conf. on Solid State Devices* (1981).
7. Saito, K. & Ito, R. *IEEE J. Quantum Electron.* **QE-16**, 205 (1980).
8. Botez, D. *IEEE J. Quantum Electron.* **QE-17**, 2290 (1981).
9. Gamo, H. *Laser Focus* **18**, 57 (1982).
10. Craighead, H.G. & Howard, R.E. *Appl. Phys. Lett.* **39**, 532 (1981).
11. Stephens, R.B. & Cody, G.D. *Solar Energy Mater.* **1**, 397 (1979).
12. Bösch, M.A. *Appl. Phys. Lett.* **40**, 8 (1982).
13. Robbins, W. *et al. Conf. on Lasers and Electro-Optics CLEO'81 Tech. Dig. 122*, Washington DC (1981).
14. Cornet, J. *et al. Conf. on Lasers and Electro-Optics CLEO'81 Tech. Dig. 122*, Washington DC (1981).
15. Bell, A.E. *J. appl. Phys.* (in the press).
16. Bell, A.E. & Spong, F.W. *IEE. J. Quantum Electron.* **QE-14**, 487 (1978).
17. Lou, D.Y. *et al. J. Vac. Sci. Technol.* **18**, 78 (1981).
18. Tu, K.N. *et al. Appl. Phys. Lett.* **39**, 927 (1981).
19. Von Gutfeld, R.J. & Chaudhari, P. *J. appl. Phys.* **43**, 4688 (1972).
20. Chen, D. *et al. IEEE Trans. Mag.* **MAG-9**, 66 (1973).
21. Chaudhari, P. *et al. Appl. Phys. Lett.* **22**, 337 (1973).
22. Bell, A.E. & Spong, F.W. *Appl. Phys. Lett.* **38** 920 (1981).
23. Smith, A.W. *Appl. Opt.* **13**, 765 (1974).
24. Chen, D. *Appl. Opt.* **13**, 767 (1974).
25. Lissberger, P.H. *J. opt. Soc. Am.* **51**, 957 (1961).
26. Mansuripur, M. *et al. SPIE Proc.* **321** (in the press).
27. Imamura, N. & Ota, C. *Conf. on Lasers and Electro-Optics CLEO'81 Tech. Dig. 106*, Washington DC (1981).
28. Connell, G.A.N. *et al. Appl. Phys. Lett.* **36**, 31 (1980).
29. Connell, G.A.N. *Appl. Phys. Lett.* **40**, 212 (1982).

Optical data storage technology is at a particularly exciting phase in its development. Steady improvements in both key areas of semiconductor laser sources as well as the optical storage medium promise the realization of optical storage memory devices within the next few years. Much further research into the microscopic degradation and ageing mechanism in thin films will be required to develop a truly archival (>100 yr) storage medium, and the development of independently addressable semiconductor laser arrays that will permit extremely high (10–100 Mbytes s^{-1}) record and playback data rates. □

Rival transmitters in visual transduction

from Helen Saibil

MANY recent papers suggest that cyclic GMP plays a major role in visual transduction, relaying the light stimulus from the visual pigment, rhodopsin, to the sodium channels of the retinal rod. Until a few years ago it was thought that this role was exclusively played by calcium ions, but the new results suggest that calcium and cyclic GMP interact to regulate the sodium conductance.

Absorption of light by rhodopsin, the visual pigment in the rod outer segment (ROS) disc membranes, changes the concentration of a cytoplasmic transmitter that controls the conductance of sodium channels in the ROS plasma membrane. These channels (or carriers) allow a continual sodium influx, or dark current, which is reduced by the absorbtion of even a single photon of light. It is this reduction (rather than an increase) of 'dark current' that brings about the neural response to a visual stimulus that is transmitted through the retina to the brain. It was originally proposed that bleached rhodopsin released stored Ca^{2+} from the discs because it was known that Ca^{2+} blocks the sodium conductance (see refs in ref 1). The recent discovery of a large and rapid light-induced release of Ca^{2+} from rods supports the idea that light causes the required rise in cytoplasmic free Ca^{2+} (refs 1,2). But recent papers on the sequence and kinetics of a series of enzyme reactions in ROS provide evidence for another mechanism in which rhodopsin bleaching controls the concentration of cGMP.

As reported in these pages about two years ago[3], rhodopsin bleaching can rapidly decrease cGMP concentration in ROS (see refs 4-6). In contrast to Ca^{2+}, injection of cGMP depolarises the rods, and delays the subsequent photoresponse[7]. Ca^{2+} could thus act as a positive transmitter, absent (or bound) in the dark and released by light to block channels[8] and cGMP could act as a negative transmitter keeping channels open in the dark, with light reducing its concentration and allowing channels to close. As well as having opposite effects on the sodium channels, Ca^{2+} and cGMP also have antagonistic effects on some phosphorylation levels in ROS[9]. Various combinations of Ca^{2+} and cGMP have been proposed to function as intracellular transmitter; these are reviewed in a recent collection of articles on the subject[10] which raises the exciting prospect that electrophysiological and biochemical approaches may be brought together from opposite ends of the transduction process.

The central question is how photoexcited rhodopsin controls the sodium conductance. From the outside of the rod, electrophysiologists have sought to characterize the sodium channels and the process controlling their conductance. From the inside, biochemists have looked for reactions activated by photoexcited rhodopsin (R*).

The first approach, using intact, functioning rods, has yielded precise information on the electrical response, showing that the transduction process has four major delay stages and that a single photoisomerization causes many channel closures (summarized in ref 11). Electrical recordings also establish that excitation and light adaptation are both mediated by diffusible transmitter substances in the outer segment[12-14]. They do not, however, provide information on the chemical nature of the controlling process, that is, the link between R* and transmitter concentration. To study this, biochemists have isolated and broken open outer segments to get at the cytoplasm where R* activation of the transduction takes place. This has led to the discovery of an enzyme cascade which links rhodopsin bleaching to cGMP hydrolysis in the ROS cytoplasm, a chain of reactions with close parallels to hormone receptor-adenylate cyclase activation[4,15].

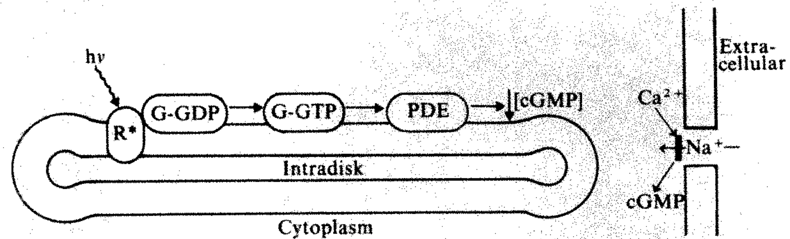
Rhodopsin bleaching stimulates this cascade, as shown in Figure 1. R* interacts with a three-subunit GTP-binding protein (called G-protein, GTPase or even transducin[15]) and catalyses its exchange of bound GDP for GTP. A single R* molecule can catalyse many such exchanges sequentially. Each G-GTP in turn activates many cGMP phosphodiesterases (PDE) which hydrolyse cGMP to 5'GMP. G and PDE, as well as a rhodopsin kinase, are bound to the cytoplasmic surface of discs[16]. G slowly hydrolyses bound GTP back to GDP, an activity which earned its original name GTPase[4], switching off PDE activation. The rhodopsin kinase may deactivate R* by ATP-dependent phosphorylation (ref.6, but see ref.9).

Changes in the far-red light scattering of ROS suspensions that were found to be elicited by a brief bleaching flash[17-19] are now being used to follow the kinetics of the earlier parts of the enzyme cascade. Kühn's key discovery that the binding of G-protein to the disc surface depends on ionic strength, light and GTP has made it possible to separate G from unbleached discs so that light scattering signals can be directly related to the amount of G present in reconstituted disc + G mixtures. Without GTP there is a strong light-induced binding of G to the discs.

Kühn, in collaboration with Bennett, Michel-Villaz and Chabre[20], went on to show that the light-induced drop in transmission (increase in turbidity) of disc suspensions in the absence of GTP reflects the formation of a 1:1 R*-G complex (fig 2). This 'binding signal' saturates when all the available G is bound (10 per cent of rhodopsin in native ROS). If > 20 μm of GTP is present, an increase in transmission, the so-called 'dissociation signal', occurs as the G-GTP dissociates from R*. This signal is seen only when GDP-GTP exchange occurs and it saturates at <0.8 per cent bleaching. Moving on to the next stage in the cascade, PDE activation, Bennett[21] used light-scattering techniques in parallel with the proton release method of Liebman and Pugh[6] to follow cGMP hydrolysis. It begins soon after the onset of the dissociation signal (Fig 2) and has the same dependence on nucleotides and light intensity, showing that PDE activation by G-GTP[15] is rapid.

The rod electrical response has a characteristic s-shaped delay which shortens with

Fig.1 Photoexcited rhodopsin (R*) binds G-protein and catalyses its exchange of bound GDP for GTP. G-GTP dissociates from R* and activates the phosphodiesterase (PDE) which hydrolyses cyclic GMP. The sodium conductance is decreased by Ca^{2+} and increased by cyclic GMP.



Helen Saibil is in the Biophysics Department, University of London King's College, Drury Lane, London WC2R 2LS.

0028-0836/82/190106-02$01.00

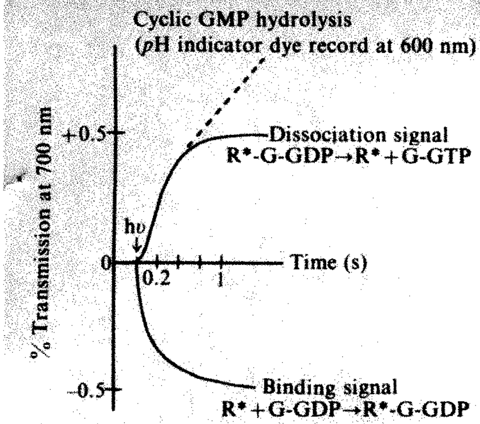


Fig. 2 Far-red light-scattering signals induced by rhodopsin bleaching in a suspension of ROS fragments or from washed discs reconstituted with isolated enzymes. The binding signal indicates the formation of a stable 1:1 complex between R* and G. The dissociation signal results from GDP–GTP exchange on G. Hydrolysis of cyclic GMP to 5′-GMP + H^+ is monitored by a pH-sensitive dye.

increasing light intensity. It is intriguing that the dissociation signal shows very similar behaviour, but it is dangerous to compare experiments on completely intact rods to those on completely disrupted ones. A possible explanation for the delay, advanced by Liebman and Pugh[6], stems from the fact that rhodopsin molecules are more abundant than the proteins with which they interact. As a result, at low bleaches when there are fewer R* created it takes longer for an isolated R* to find its target protein — a process dependent on lateral diffusion. Lateral diffusion seems rate-limiting in the analogous hormone receptor system where chemically increasing the membrane fluidity speeds up cyclase activation[22].

Both Ca^{2+} and cGMP have drawbacks as candidates for the intracellular transmitter. From the biochemical viewpoint, the persistent failure to relate Ca^{2+} uptake or release to changes in rhodopsin or disc membranes[23] and the relative completeness of the R*-cGMP pathway makes cGMP an attractive alternative. From the electrophysiological viewpoint, cGMP hydrolysis has kinetics comparable to the electrical response only if the hydrolysis is triggered by light levels much higher than needed to trigger the corresponding electrical response (see refs in refs 5 & 11), and there are questions about its time course *in vivo*[24]. We have had already had to accept that the minor proteins in ROS cannot be dismissed as irrelevant to visual excitation; now we may have to learn to live with multiple regulation of the sodium conductance by calcium and cGMP. □

1. Yoshikami, S., George, J.S. & Hagins, W.A. *Nature* **286**, 395 (1980).
2. Gold, G.H. & Korenbrot, J.I. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5557 980).
3. Fatt, P. *Nature* **280**, 355 (1979).
4. Bitensky, M.W., Wheeler, G.L., Yamazaki, A., Rasenick, M.M. & Stein, P.J. in ref. 10, p.238.
5. Bownds, M.D. in ref. 10, p.203.
6. Liebman, P.A. & Pugh, E.N. in ref. 10, p.157.
7. Miller, W.H. & Nicol, G.D. in ref. 10, p.417.
8. Yau, K.-W., McNaughton, P.A. & Hodgkin, A.L. *Nature* **292**, 502 (1981).
9. Hermolin, J., Karell, M.A., Hamm, H.E. & Bownds, M.D. *J. gen. Physiol.* **79**, 633 (1982).
10. Miller, W.H. (ed.) *Curr. Topics in Membranes and Transport* **15** (1981).
11. Matthews, G. & Baylor, D.A. in ref. 10, p.4.
12. Donner, K.O. & Hemilä, S. *Med. Biol.* **56**, 52 (1978).
13. Bastian, B.L. & Fain, G.L. in ref. 10, p.341.
14. Lamb, T.D., McNaughton, P.A. & Yau, K.-W. *J. Physiol., Lond.* **319**, 463 (1981).
15. Fung, B.K.-K., Hurley, J.B. & Stryer, L. *Proc. natn. Acad. Sci. U.S.A.* **78**, 152 (1981).
16. Kühn, H. in ref. 10, p.172.
17. Asai, H., Chiba, T., Kimura, S. & Takagi, M. *Expl Eye Res.* **21**, 259 (1975).
18. Bignetti, E., Cavaggioni, A., Fasella, P., Ottonello, S. & Rossi, G.L. *Molec. cell. Biochem.* **30**, 93 (1980).
19. Reichert, J. & Hofmann, K.P. *Biophys. struct. Mech.* **8**, 95 (1981).
20. Kühn, H., Bennett, N., Michel-Villaz, M. & Chabre, M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6873 (1981).
21. Bennett, N. *Eur. J. Biochem.* **123**, 133 (1982).
22. Hanski, E., Rimon, G. & Levitzki, A. *Biochemistry* **18**, 846 (1979).
23. Szuts, E. in ref. 10, p.291.
24. Ebrey, T.G., Kilbride, P., Hurley, J.B., Calhoon, R. & Tsuda, M. in ref. 10, p.121.

Scrapie agent: prions or virinos?

from Richard H. Kimberlin

In the last few weeks considerable publicity has been given to a claim, by Stanley B. Prusiner of the University of California, San Francisco, that the agent causing scrapie — a degenerative disease of the central nervous system in sheep and goats — is a novel kind of proteinaceous infectious particle[1]. The particle, named the 'prion' by Prusiner, is claimed to be either a very small oligonucleotide surrounded by a tightly packed protein or, more remarkably, an infectious protein that contains no nucleic acid at all. This latter possibility is, as Prusiner says, 'clearly heretical' and raises all the problems of how a non-nucleic acid-containing particle can replicate itself. Have we indeed reached the point where the existence of an infectious protein seems likely? To answer this question Prusiner's work needs to be set in the context of 20 years' research in a very difficult area.

In the early 1960s, scrapie was experimentally transmitted from its natural hosts, sheep and goats, to mice and other rodent species, allowing a wide variety of experimental models of scrapie, novel techniques for identifying different strains of agent and a conceptual framework for the genetic control and slow pathogenesis of the disease. Equally important, quantitative studies on the nature of the transmissible agent could begin using titration in mice to measure infectivity (see ref. 2 for a review).

Since then, various classes of microbial agents, such as bacteria, mycoplasmas and viroids, have been excluded as the scrapie agent. The agent has many of the biological properties of a virus but is 'unconventional' in its high degree of physicochemical stability and apparent lack of specific antigenicity. However, we still have very little detailed knowledge of the nature of the scrapie agent.

There are two main reasons for this slow progress. First, the infectious agent is notoriously 'sticky': it seems to aggregate to cell components and has not been sufficiently purified to make biochemical characterization easy. Consequently, its physicochemical properties have been determined only on crude or partially purified extracts. Second, titration in animals is still the only method of measuring the infectious agent. Even with the quickest models of scrapie in mice and hamsters, a complete titration takes 200–300 days. There are also difficulties in interpreting the results: infectivity titres obviously reflect the amount of scrapie agent in an inoculum but they are also influenced by the efficiency of the infection process. This in turn depends on factors which may include varying host constituents and residual reagents in inocula, and the degree of dispersion or aggregation of a 'sticky' agent after treatment. A change in scrapie titre may be particularly difficult to interpret if a given reagent interacts with more than one molecular species, for example, protein and nucleic acid.

Prusiner has worked largely with a short-incubation hamster model of scrapie and has added a sizeable amount of evidence, supporting earlier observations that scrapie infectivity can be reduced by treatments which denature, disaggregate or breakdown proteins[3-6]. The loss of infectivity is generally quite large and it seems reasonable to conclude that the infectivity of scrapie agent involves protein.

However, the unresolved issue is whether such protein is 'contained' in the agent as an integral part of its structure or merely associated with it. This leads to the key question: is there a scrapie-specific protein? Neither Prusiner nor anyone else has any published data showing that there is, but, for the moment, let us assume that one does exist. Conventional thinking would then be that the protein is encoded by a specific nucleic acid which forms the genome of the agent. However, Prusiner has produced two interesting lines of evidence which he regards as indicating the absence of a scrapie nucleic acid.

First, treatments known to inactivate nucleic acids, even though not specific for

Richard H. Kimberlin is Head of the Pathogenesis Section, ARC & MRC Neuropathogenesis Unit, Edinburgh.

0028-0835/82/190107-02$01.00

them, are not found to destroy scrapie infectivity. Infectivity is relatively resistant to heating at 65°C in the presence of zinc ions, to photochemical inactivation after treatment with psoralens and to treatment with 0.5 M hydroxylamine[1]. Failure to inactivate may, however, simply mean that these reagents did not gain access to the putative nucleic acid, perhaps because of protection by protein forming part of the infectious agent or by contaminating host molecules. In any case, as Prusiner points out, not all viruses are inactivated by psoralens or hydroxylamine.

Second, in preliminary studies using detergents and gel filtration, some scrapie infectivity eluted between bovine serum albumin and ovalbumin markers[1]. Assuming a globular structure for the infectious agent, and recognizing the potential problems of determining molecular weights in this way, Prusiner argues that scrapie agent may have a molecular weight of 50,000 or less. He considers that a proteinaceous agent of this size would be too small to contain a nucleic acid of more than just a few nucleotides.

Interesting though they are, these two findings are not by themselves compelling reasons for considering highly unorthodox models of scrapie agent, particularly when faced by some of the problems these pose. Prusiner suggests three such models.

In the first model, the agent contains no nucleic acid and scrapie-specific protein acts as its own template, either directly, or by 'reverse translation'. It is difficult to conceive how proteins could code for their own synthesis although scrapie models of this kind were proposed as long ago as 1967 (ref. 7). But since then the requirements of a scrapie model have become more stringent; the copying process must have a precision compatible with the existence of several different, genetically stable strains of scrapie[8].

In Prusiner's second model, the agent contains no nucleic acid but has protein which induces host genes to code for its own synthesis. The third proposes that there is a very small scrapie oligonucleotide which does not code for protein but acts as a regulatory element promoting the synthesis of agent. In these models, therefore, it is the host nucleic acid that encodes scrapie-specific information. This is certainly conceivable but difficult to reconcile with evidence that scrapie agents behave like pathogens with their own coded information: the natural disease is infectious[9], various strains of agent can be transmitted experimentally to many different species[8], some strains give rise to mutants[10] and strain selection can occur on serial passage of mixtures[11]. To explain these properties in terms of host coding introduces a complexity that seems unnecessary in the absence of data which necessitate it.

A much simpler working hypothesis which fits both established facts and even Prusiner's recent data may be proposed. It makes only two assumptions: first, that there is indeed a scrapie-specific nucleic acid, as indicated by the properties of replication, strain variation and mutation; and second, that, as in viroids, the scrapie nucleic acid is not translated[12].

The hypothesis has some interesting consequences. The scrapie nucleic acid could be very small as all it has to do is be replicated (using host enzymes) and interact in strain-specific ways with the host to produce disease. The latter it could do by acting as, or coding for, some kind of regulatory nucleic acid. Alternatively, disease could be a consequence of the binding of the scrapie-specific nucleic acid to host protein needed to form an infectious agent, especially if the protein normally has an important biological function. Since the protein associated with the agent is host derived we have a very simple explanation for the apparent absence of specific antigenicity.

This hypothesis may also explain our current failure to purify agent or to identify it in the electron microscope: if the scrapie-specific nucleic acid is covered with host protein it may well be 'sticky' and hard to purify from other host proteins and, unless it has a regular, virus-like nucleoprotein structure, it could be hard to recognize. An agent of this kind would be novel but fit nicely into a biological niche lying between viruses, which specify some of their own proteins, and viroids which need no protein at all. The neologism, 'virino', has already been suggested for a novel agent of this kind[13], and may by preferable to Prusiner's term 'prion' because the latter emphasizes a molecular species which may not be the most important one, as did the conclusion in 1935 that tobacco mosaic virus was proteinaceous[14].

However, the main point of these speculations is that we do not yet need to build hypotheses outside the current framework of molecular biology to accommodate the scrapie agent. The real need is to assemble more hard facts and, just as Prusiner[1] says, this means "purification of the scrapie agent to homogeneity and the subsequent identification of its macromolecular components"; undoubtedly a difficult task but not an impossible one. □

1. Prusiner, S. B. *Science* **216**, 136 (1982).
2. Kimberlin, R. H. in *Comparative Diagnosis of Viral Diseases* Vol. 3A (eds Kurstak, E. & Kurstak, C.) 349 (Academic, New York, 1981).
3. Hunter, G. D. in *Slow Transmissible Diseases of the Nervous System* Vol. 2 (eds Prusiner, S. B. & Hadlow, W. J.) 365 (Academic, New York, 1979).
4. Latarjet, R. in *Slow Transmissible Diseases of the Nervous System* Vol. 2 (eds Prusiner, S. B. & Hadlow, W. J.) 387 (Academic, New York, 1979).
5. Millson, G. C. & Manning, E. J. in *Slow Transmissible Diseases of the Nervous System* Vol. 2 (eds Prusiner, S. B. & Hadlow, W. J.) 409 (Academic, New York, 1979).
6. Rohwer, R. G., Brown, P. W. & Gajdusek, D. C. in *Slow Transmissible Diseases of the Nervous System* Vol. 2 (eds Prusiner, S. B. & Hadlow, W. J.) 456 (Academic, New York, 1979).
7. Griffiths, J. S. *Nature* **215**, 1043 (1967).
8. Dickinson, A. G. & Fraser, H. in *Slow Transmissible Diseases of the Nervous System* Vol. 1 (eds Prusiner S. B. & Hadlow, W. J.) 367 (Academic, New York, 1979).
9. Kimberlin, R. H. *Nature* **278**, 303 (1979).
10. Bruce, M. E. & Dickinson, A. G. in *Slow Transmissible Diseases of the Nervous System* Vol. 2 (eds Prusiner, S. B. & Hadlow, W. J.) 71 (Academic, New York, 1979).
11. Kimberlin, R. H. & Walker, C. A. *J. gen. Virol.* **37**, 487 (1978).
12. Diener, T. O. *Viroids and Viroid Diseases* (Wiley, New York, 1979).
13. Dickinson, A. G. & Outram, G. W. in *Slow Transmissible Diseases of the Nervous System* Vol. 2 (eds Prusiner, S. B. & Hadlow, W. J.) 13 (Academic, New York, 1979).
14. Stanley, W. M. *Science* **81**, 644 (1935).

NATURE

100 years ago

THAT the French should know better than any other nation how to enlist art in the service of science is just what might be expected. Such a service on the part of art to science is a fair return for the immense resources which scientific research has been able to place at the disposal of art. Nowhere have the discoveries of science been more useful or more utilised than at the celebrated porcelain manufactory of Sèvres, and the illustrations which we give to day will afford some idea of the beautiful results which are thus produced. As a permanent record of successful scientific efforts, nothing could be more satisfactory and appropriate. In the figure the characteristic features of the Arctic regions are rendered with almost perfect success and truthfulness.

From *Nature* **26**, 36, 11 May 1882.

Sèvres Vase, commemorative of the North-East Voyage of Baron Nordenskjöld.

REVIEW ARTICLE

Aspects of the electrochemistry of steel in concrete

C. L. Page

Department of Construction and Environmental Health, University of Aston, Gosta Green, Birmingham B4 7ET, UK

K. W. J. Treadaway

Building Research Establishment, Garston, Watford WD2 7JR, UK

Information available on mechanisms of corrosion of steel in concrete is reviewed and the need emphasized for a better understanding of the electrochemistry of the subject and aspects of the physical performance of the concerete cover in relation to the diffusion of both chloride and oxygen. The practical influence of data obtained from studies in such areas in relation to longer-term improvements in durability of reinforced concrete is considered.

THE practice of embedding steel bars in a mixture of Portland cement, graded aggregates and water to form reinforced concrete was introduced during the second half of the nineteenth century. Initially the material was employed for projects of modest scale but the scope of its applications grew rapidly as its composite properties were appreciated.

Since the turn of the present century reinforced concrete has been in widespread use in a great variety of building and civil engineering work. It is exposed to many different types of environment and, as for any structural materials, the related characteristics of durability and need for maintenance are of considerable importance as service lives are expected to extend over several decades. In these respects, the performance of reinforced concrete has generally been impressive and there are many examples of its long-term durability (see ref. 1).

There are, however, several degradative processes which affect a minority of reinforced concrete structures leading to loss of serviceability or, in extreme cases, to structural collapse and, amongst these, the most common cause of deterioration is corrosion of reinforcing steel[2]. This gives rise to a familiar sequence of events owing to the volume increase associated with the transformation of steel to rust which generates tensile stress within the surrounding concrete and causes cracking and eventual spalling of the cover to the reinforcement. Some examples of the types of problems and potential safety hazards which reinforcement corrosion can cause in buildings are illustrated in Fig. 1. Similar corrosion problems are also of major importance worldwide in relation to reinforced concrete road bridges and marine structures.

Despite the length of time during which experience has been gained of the corrosion behaviour of steel in concrete, there are still several important aspects that are not fully understood: in particular, the interpretation of information obtained from relatively short-term electrochemical studies, which are frequently used as an aid in predicting long-term performance[3]. Because such investigations often represent the only practicable approach for rapid assessment of the effects on corrosion behaviour of changes in materials specifications, concreting practices and so on, it is important to seek improved fundamental understanding of the electrochemistry of this complex and variable system.

We now survey briefly the commonly encountered causes of reinforcement corrosion and highlight some main areas where basic research is considered necessary.

Mechanisms of corrosion protection in concrete

Any attempt to describe in general terms the nature of concrete, viewed as a potentially corrosive medium, is likely to prove inadequate because many factors can affect both the physical pore structure of the material and the chemical composition of the solution phase within the pores. The most important constituent is the hydraulic binder, Portland cement, which, in its unhydrated form, consists mainly of four minerals of the following approximate compositions: tricalcium silicate ($3CaO\,SiO_2$), dicalcium silicate ($2CaO\,SiO_2$), tricalcium aluminate ($3CaO\,Al_2O_3$), tetracalcium alumino-ferrite ($4CaO\,Al_2O_3\,Fe_2O_3$) (In cement chemistry these minerals are conventionally referred to as C_3S, C_2S, C_3A and C_4AF.) When Portland cement is mixed with water, the constituent minerals undergo a complex sequence of hydration reactions, which gradually transform the paste to a hardened matrix of hydrated products. The most important reactions, insofar as the development of physical structure and mechanical properties are concerned, are those involving C_3S and C_2S which are converted to a gel of calcium silicate hydrates of ill-defined composition and structure (C–S–H gel) and to calcium hydroxide produced mainly in the form of well-developed portlandite crystals[4]. In a mature (fully hydrated) sample, C–S–H gel would be expected to account for ~70% of the weight of solid material and calcium hydroxide for ~20% (ref. 5), the eventual porosity and pore size distribution being determined largely by the water/cement ratio, as illustrated in Fig. 2.

From the point of view of the electrochemical behaviour of steel in concrete, the most significant feature of cement hydration is that the aqueous phase rapidly acquires a high pH value. Initially, the solution in contact with the hydrating cement grains contains hydroxides and sulphates of calcium, sodium and potassium but sulphate ions are rapidly precipitated to form the highly insoluble calcium sulpho-aluminate hydrate ($C_3A\,3CaSO_4\,31H_2O$) (ref. 6). Thereafter, the liquid becomes increasingly concentrated with respect to sodium and potassium hydroxides and, as the pH rises, the concentration of calcium ions decreases in approximate conformity with the solubility product of calcium hydroxide. This state of affairs is reflected in the fact that, after a few weeks of hydration, pH values well in excess of 13 are frequently recorded for Portland cement pastes of water/cement ratios such as are commonly used in practice[7-9].

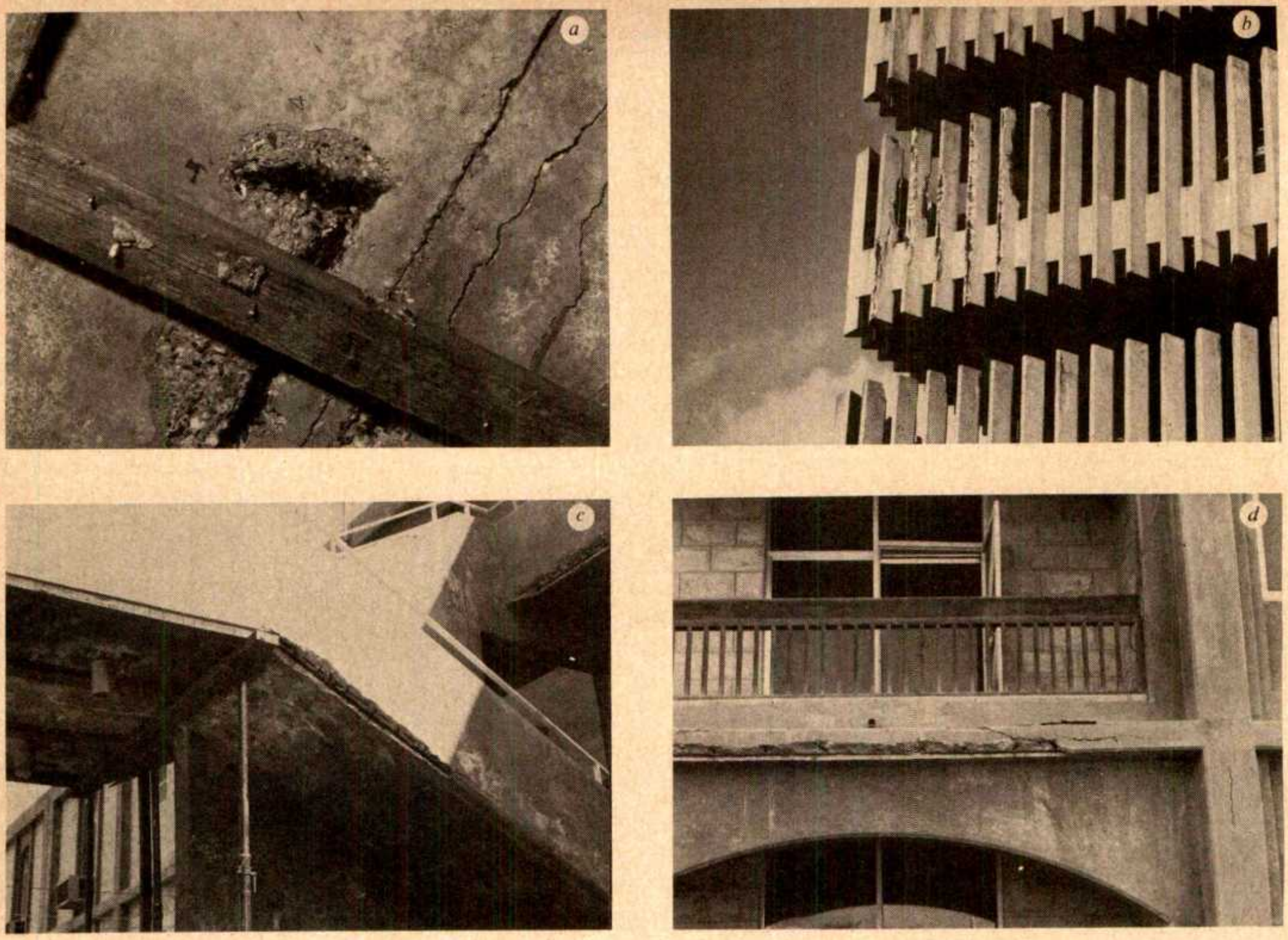


Fig. 1 *a*, Cracking and spalling in the soffit of precast concrete roof planks resulting from reinforcement corrosion and leading to a loss of serviceability requiring roof replacement. *b*, Corrosion revealed when cracked concrete cover was removed from a number of precast concrete fins in a multi-storey car park. *c*, Propping of a staircase necessitated by structural weakening as a result of reinforcement corrosion (in a Middle Eastern structure). *d*, Spalling of cover to reinforcement in beams, and cracking of columns, as a result of reinforcement corrosion in a Middle Eastern building.

For iron in an alkaline environment, the Pourbaix diagram indicates that the metal may remain passive over a wide range or potentials[10] and a very limited supply of oxygen is sufficient to provide the necessary degree of anodic polarization to maintain conditions within this range, as shown by rest potential measurement in deaerated limewater[11]. Thus, in all but highly anaerobic environments, the primary mechanism of corrosion protection in concrete involves passivation of embedded steel resulting from the chemically inhibitive influence of hydroxyl ions.

The secondary feature of corrosion protection that is provided by dense concrete depends on the capacity of the material to act as a physical barrier between embedded reinforcement and the external environment, limiting the access of substances which may play a part in corrosion. In this respect, however, it is clear that concrete, with its continuous pore system and tendency to form surface cracks, is a far from perfect barrier. Thus concrete is not generally effective in excluding water but this may not cause corrosion problems unless the pH falls below that required to maintain passivity. Where the pH does fall, however, corrosion problems can arise even when the material is exposed to the air inside buildings provided the appropriate circumstances prevail[12]. It has also been established that dense concrete is normally fairly permeable to oxygen and, although the rate of oxygen diffusion is obviously limited by increase in the degree of saturation of the material, a significant flux can be sustained even through specimens that are fully submerged for long periods in water[13]. The real importance of the physical barrier presented by concrete cover is therefore related mainly to its ability to preserve the conditions of high pH needed to maintain the reinforcement in a passive condition by limiting the rate of ingress of acidic substances from the external environment.

Generally the acidic substances concerned are atmospheric carbon dioxide and, in polluted locations, other gases such as sulphur dioxide. These react with the alkaline constituents of the cement paste to form a carbonated zone which gradually penetrates into exposed concrete, reducing the pH of the affected region to a value below pH9.

Loss of protection by carbonation

The kinetics of carbonation for various types of concrete in different environments have been widely investigated and it has generally been observed[14] that the relationship between carbonation layer thickness (x) and exposure time (t) is approximately parabolic in form:

$$\frac{dx}{dt} = kt^{-1/2}$$

The magnitude of the proportionality constant (k) depends on several variables related to the quality of the concrete (water/cement ratio, type of cement and so on) and to the environment (temperature, relative humidity and so on). By specifying a suitable grade of concrete for a given service situation, it is possible to ensure that the rate of advance of

carbonation dx/dt declines within a fairly short time to a low level (<1 mm yr^{-1}) Provided the depth of concrete cover to the reinforcement is adequate, therefore, carbonation should not penetrate to an extent which endangers the passivity of the steel during the design life of a structure It is this principle that underlies recommendations made in codes of practice regarding minimum depths of cover appropriate for concretes of different quality exposed to environments of varying severity[15]

It is particularly important when concrete is being produced in adverse climatic conditions that adequate precautions be taken to ensure proper curing of the external surfaces because the formation of a weak, permeable surface layer would otherwise tend to promote carbonation[16] In general, however, problems of corrosion arising from carbonation should be avoidable in normal density structural concrete, provided that well-established guidelines for constructional practice are followed The situation that most frequently gives rise to corrosion risks, altogether more serious and difficult to contend with, occurs when reinforced concrete contains a substantial amount of chloride ion as this may cause depassivation of steel even in a medium of high pH such as uncarbonated concrete

Loss of protection by chloride

Chlorides may be introduced into concrete during manufacture or service Within the former category, there is the posibility of deliberate inclusion of admixtures containing calcium chloride to accelerate the early stages of cement hydration, a practice that is now strongly discouraged for reinforced concrete[17] There are also many cases of adventitious introduction of chlorides as contaminants of the aggregates or mixing water, which may be almost unavoidable in certain circumstances where the use of locally available resources is necessary[18,19] During service, chlorides from external sources may penetrate concrete to levels which exceed normal depths of cover to reinforcement within a small fraction of the design life in the case of structures exposed to marine conditions[20,21] or to de-icing salts[22,23]

Several matters arising in relation to the behaviour of steel in chloride-bearing concretes remain incompletely resolved despite long and numerous researches, the corrosive effects of calcium chloride in concrete were, in fact, studied as long ago as 1923 (ref 24) We now consider some of the outstanding difficulties

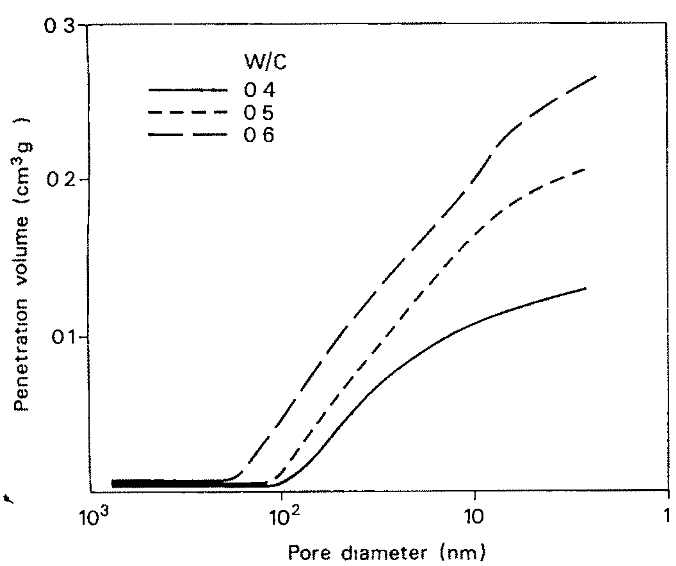


Fig. 2 Pore size distribution curves for mature hydrated ordinary Portland cement pastes determined by mercury intrusion porosimetry (Penetration volume indicates cumulative pore volume intruded by mercury, W/C indicates water cement ratio of pastes)

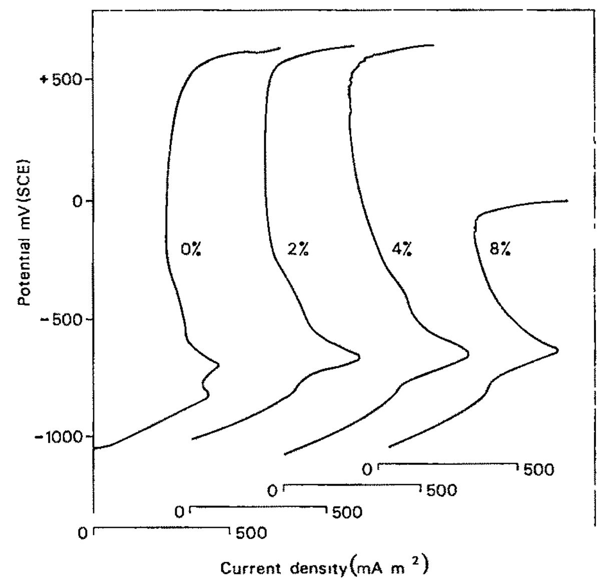


Fig. 3 Potentiodynamic anodic polarization of mild steel in hydrated ordinary Portland cement pastes of water/cement ratio 0 4 containing various percentages of calcium chloride dihydrate by weight of cement (sweep rate = 1 mV s^{-1})

A question of practical importance, that has to be considered when chloride-bearing concreting materials are specified or when investigations of the durability of reinforced concrete structures are made, concerns the threshold level of chloride ion which is liable to stimulate corrosion of steel in concrete As might reasonably be expected, however, bearing in mind the range of possible variations in concrete quality and exposure conditions, there is no simple answer to this question and on the basis of surveys of large numbers of reinforced concrete structures in Britain, the Building Research Establishment has recently proposed a classification for assessing the risk of corrosion[25] Thus for concrete made from ordinary Portland cement (or similar), it is considered that a low risk is associated with chloride contents (by weight of the cement) of <0 4%, an intermediate risk for chloride contents of 0 4–1 0% and a high risk for levels >1 0% The influence of carbonation in this process is also considered, the risk being enhanced when carbonation reaches the reinforcement

Whilst providing sound practical guidelines that are adequate for many purposes, the above classification is not based on any mechanistic interpretations of the factors which determine the risk of corrosion in concrete containing a given quantity of chloride ion It is therefore limiting in the sense that it provides no means for distinguishing *a priori* between the effects of a number of possible sources of variation, for example, the following (1) chlorides associated with different cations, (2) chlorides introduced at or after the time of mixing, (3) chlorides introduced in the presence of other contaminant anions such as sulphates, (4) chlorides in concrete made from cements of different compositions, (5) chlorides in concrete of varying water/cement ratio, (6) chlorides in concrete cured or exposed in different conditions

Because these and other factors are likely to influence the corrosion risk associated with the presence of a given concentration of chloride ions, there is a need to examine their roles and relative importance

Mechanisms of corrosion of steel in concrete

The simplest and most reliable method involves long-term exposure tests on laboratory-prepared specimens of reinforced concrete, a costly and time-consuming procedure but which yields results that can generally be considered as a realistic guide to practical performance The Building Research Establishment have used this approach in studying the influence of

many factors, such as the use of various types of reinforcing steel, and found that differences in performance may take several years to manifest themselves clearly[3]

The alternative methods, that are often used when information has to be obtained within a restricted time, involve the application of various electrochemical techniques They may be applied to steel which is either embedded in concrete, mortar, cement paste and so on, or else immersed in a solution of which the composition is believed to resemble that of the pore electrolyte in the concrete under consideration With all of these methods, there are considerable difficulties in relating the results obtained to the performance of reinforced concrete as recorded from service experience or from exposure tests

In studies involving the use of simulated pore-solutions there are fairly obvious uncertainties regarding the nature and concentrations of solutes that should be present Thus it has commonly been assumed that saturated calcium hydroxide solution, which has a pH value of ~12 6 at 25 °C, may be taken as an approximate electrolyte for electrochemical studies This, however, is apparently at variance with the findings of several investigations of the composition of pore solutions extracted from actual Portland cement pastes or mortars, these have shown that, after a few weeks of curing, the pore liquid has often become essentially a mixed solution of sodium hydroxide and potassium hydroxide with a pH in the range 13–14 (refs 7–9) The latter electrolyte would be expected to maintain the passivity of steel in the presence of chloride concentrations well in excess of the level (~500–700 p p m) which induces depassivation in saturated calcium hydroxide[11 26]

Before concluding that a more alkaline solution would constitute a more realistic model, however, note that the pore-solution in concrete is not buffered at a pH value in excess of ~pH 12 6 and several commonplace processes may tend to reduce the naturally high alkalinities produced by cements rich in sodium and potassium oxides Examples of such processes are (1) use of calcium salts, for example, calcium chloride, as admixtures which would lead to precipitation of hydroxyl ions to form additional calcium hydroxide, (2) ingress of cations which form insoluble hydroxides from certain environments, for example magnesium ion from seawater, (3) leaching of sodium and potassium hydroxide from concrete frequently or continuously exposed to water, (4) diffusion of hydroxyl ion from the interior towards the carbonated surface zone of concrete exposed to the atmosphere, (5) simple dilution of the pore-solution associated with the cement paste by water from continuous pores within the aggregates

It is perhaps reasonable, therefore, to consider that, in service conditions, the pH of uncarbonated concrete at the level of reinforcing bars may often tend towards the buffering range of saturated calcium hydroxide[27] This suggestion would fit in with the otherwise surprising observation that cements with high alkali-contents have not been found to provide significantly enhanced corrosion protection to steel in reinforced concrete exposed for long periods to chloride-bearing environments[28]

Apart from the question of pH, the other principal uncertainty regarding the composition of the liquid phase in concrete concerns the extent of binding of the chloride ions by solid hydration products It is well-known that the aluminate and ferrite constituents of Portland cements can form incongruently soluble calcium chloro-aluminate and chloro-ferrite hydrates the former being the more important from the point of view of their capacity to remove free chloride ions from solution[29 30] There is also considerable evidence that the C_3A content of the cement has an important role in determining the extent to which chlorides included in the mix constituents are complexed and that this influences the degree of protection afforded to embedded steel in the presence of chloride[29,31 32], this is recognized in codes of practice by the imposition of stringent controls over the use of chloride-containing materials in concretes manufactured from cements of specified low C_3A content, such as sulphate resisting Portland cement[17]

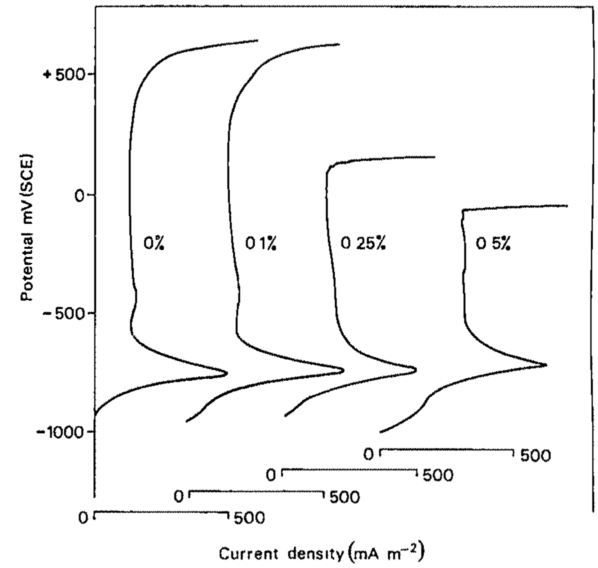


Fig. 4 Potentiodynamic anodic polarization of mild steel in saturated calcium hydroxide containing various percentages of calcium chloride dihydrate (sweep rate = 1 mV s^{-1})

Nevertheless, there is a general lack of information about the levels of chloride likely to remain uncombined for long periods in the solution phase associated with concretes made from various cements in different conditions of natural exposure The source of the chloride ions may be significant and it has been suggested that chlorides which penetrate concrete from the external environment tend to remain uncomplexed to a greater degree than those present at the time of manufacture[21 33] It is also unclear to what extent the presence of various levels of sulphate ions, either as constituents of the mix materials or as contaminants entering from the service environment, may liberate chloride ions owing to preferential formation of calcium sulpho-aluminate hydrates Similarly the extent to which carbonation may cause decomposition of calcium chloro-aluminate hydrates is uncertain

The above considerations highlight the need for further investigations of the solution chemistry of concrete pore electrolytes, aimed particularly at resolving the role of cement compositional variables (such as C_3A content and alkali content) in relation to several of the possible long-term changes that have been identified Examination of solutions extracted from concretes, exposed for long periods to different service environments, is practicable and would appear to be an important field for future work

Whilst improved understanding of the constitution of the aqueous phase within concrete would greatly facilitate attempts to understand the electrochemistry of steel in concrete by means of experiments involving steel electrodes immersed in simulated pore-solutions, the latter can never be an entirely satisfactory model The main objection is that the kinetics of diffusion of the reactants and products of electrode processes cannot be comparable in solutions and in concrete

As has been noted previously[16 27], depassivation as a result of chloride-induced pitting of steel in concrete is likely to be controlled by the rates of diffusion of several substances within the material One significant factor is the rate of oxygen transport to the metal surface as this will influence cathodic polarization and determine whether the potential of the steel is sufficiently noble to support the formation and growth of pits at a given effective concentration of chloride and pH (ref 34) For reinforced concrete, fully submerged in a chloride-bearing aqueous environment, pitting is often likely to be suppressed[27] and the only question that arises is whether the supply of oxygen to the steel will be even sufficient to balance the very low anodic leakage current density required to maintain the passive film[35] or whether a low-potential 'active' condition will result[36 37] It

has also been suggested that the magnitude of this leakage current density is itself controlled by diffusional constraints on the passage of iron ions into the surrounding medium, and may vary in response to change in the surface concentrations of chloride and hydroxyl ions[38]. In conditions other than those of total immersion, the supply of oxygen to embedded steel is unlikely to be a limiting factor and the availability and rates of replenishment of chloride and hydroxyl ions seem liable to govern the anodic behaviour of the metal. Notwithstanding the controversy regarding mechanisms of chloride-induced pit initiation, it is generally accepted that pit growth is accompanied by the development of solution conditions within the pits of locally elevated chloride ion concentration and depressed *p*H (ref 34). Factors which limit the diffusion of chloride ions or enhance the diffusion of hydroxyl ions in the vicinity of incipient pits will therefore tend to promote repassivation. One such factor is associated with the composition and physical barrier-layer characteristics of the lime-rich interfacial zone of segregation which surrounds steel embedded in cement-rich dense concrete[16,27], others are related to the mineralogy and pore structure of the hydrated cement matrix and, in this regard, it is significant that effective diffusivities of chloride ion, measured at fixed temperature for different cement pastes of constant water/cement ratio, have been found to vary widely[39,40]. Activation energies recorded for chloride ion diffusion in hydrated cement pastes indicate that the kinetics are governed by interactions between the diffusing ions and the pore surfaces[39,40]. The nature of these interactions and the effects on them of cement compositional variables represent a potentially useful field of continuing research. This may have practical applications related to the development of hydraulic cements which offer a high degree of corrosion protection to embedded steel by virtue of their capacity to retard the diffusion of chloride ions[41,42].

The foregoing discussion has indicated several major shortcomings of electrochemical investigations performed with electrodes immersed in solutions. To simulate the various dynamic factors which may play a part in controlling the durability of reinforcement in concrete, it is therefore necessary to undertake studies with electrodes embedded in concrete, mortar or cement pastes and several attempts have been made to carry out electrochemical measurements with specimens of these types. As will be discussed, the results obtained must generally be interpreted with considerable caution.

One of the greatest practical problems associated with all such investigations concerns the method of masking the embedded electrode so that it may be connected to external apparatus with a predetermined part of its surface area exposed to the electrolytic medium. The avoidance of crevice effects in the specimens is very difficult and the situation is exacerbated by the fact that many commonly used lacquers, coatings and gasket materials are themselves likely to suffer some degree of degradation during exposure for long periods to a highly alkaline environment such as concrete. When quantitative data are sought from long-term electrochemical studies of this nature, specimen design is therefore crucial if experimental artefacts are to be avoided. In particular, we suggest that, in planning such experiments it is worthwhile to consider using a range of specimens with systematically varied exposed surface area as this may facilitate the subsequent diagnosis of misleading effects of the type mentioned.

In the case of standard methods, such as potentiodynamic or potentiostatic polarization curve determinations, the data obtained will depend on variables such as the rate of change of potential and they have little prospect of yielding information which can be directly related to long-term durability. Their value is purely comparative in the sense that they can be used to provide rapid evidence of trends associated with a particular variable which is altered whilst others are held constant. Thus, for example, using fixed experimental conditions, different types of alloy steel may be graded according to their relative susceptibilities towards pitting in particular chloride-containing mortars[3], or the relative effects of given dosages of chloride may be examined for steel embedded in various types of cement pastes and mortars[43–45].

As the typical data shown in Fig 3 illustrate, however, the levels of chloride that are observed to cause significant effects in short-term experiments of this type are often considerably greater than those which are associated with a high risk of corrosion in long-term exposure tests or service. Note that the concentration of chloride ions in the original mixing water of the paste specimen containing 4% of calcium chloride dihydrate by weight of the cement, which appears to be on the borderline between passivity and pitting (see Fig 3), is almost 1,000 times larger than the concentration required to induce pitting in saturated calcium hydroxide solution (see Fig 4). This scale of disparity cannot be adequately explained in terms of the effects of chloro-aluminate hydrate formation or the excess hydroxyl ions contributed by sodium and potassium oxides in the cement, particularly bearing in mind the *p*H-depressing action of calcium chloride[44]. The main factor is probably the slowness of chloride diffusion towards the embedded steel surfaces, which causes pitting to be 'missed' even at a relatively low rate of potential scan unless very large chloride concentrations are present in the pore solution.

Similar considerations apply, in general, to the use of galvanostatic polarization for examining the response of embedded steel electrodes. The technique has been useful as a means of providing comparative data[3,46–50]. There is little doubt, however, that whilst this approach may be more reliable than potentiostatic or potentiodynamic methods when small current densities are applied, concentration polarization distorts anodic behaviour to an extent which makes accurate prediction of long-term performance impossible. A further disadvantage, common to galvanostatic, potentiodynamic and potentiostatic polarization curve determinations, is that the specimens used may be examined once only because of the likelihood of irreversible changes being brought about by potential attenuation over a range of several hundred millivolts, this effectively eliminates the possibility of following time-dependent changes in the response of individual specimens.

There are various alternative electrochemical methods which have been applied to steel in concrete or similar media in attempts to provide data continuously or at timed intervals over a period of several months or more. The range of techniques used has included rest potential measurements[50,51], current measurement in potentiostatic conditions in deaerated media[35], polarization resistance determination[52,53], and a c impedance measurement[54,55]. In addition to the practical difficulties mentioned regarding electrode design to avoid misleading crevice effects and so on there are other grounds for caution in interpreting data from long-term electrochemical investigations carried out in concrete and, even in straightforward rest-potential studies, ambiguities can exist as other workers have

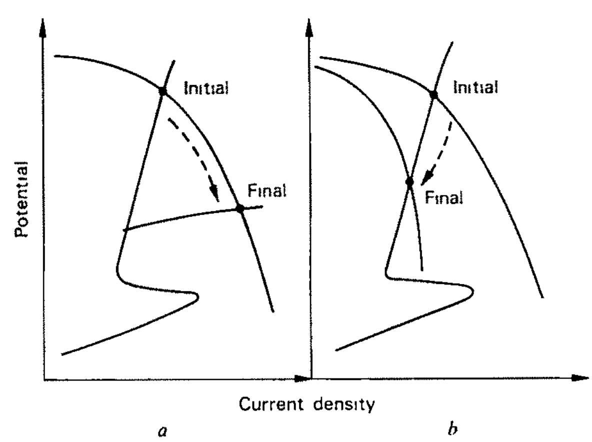


Fig. 5 Movements in rest-potential towards more negative values associated with *a*, passive film breakdown, *b*, restriction of oxygen access

indicated[36 50] Thus, it is well-established that passivity is frequently associated with a noble potential, which tends to increase with time towards a steady value, usually in excess of about −220 mV SCE (ref 51) as the passive film thickens, implying a plentiful supply of oxygen In cases where the rest-potential is observed to decrease with time, depassivation may or may not be implied since the increased cathodic polarization may be associated either with passive film breakdown (Fig 5a) or with progressive restriction in the supply of oxygen to the surface of passive steel (Fig 5b)

Of the other techniques referred to, it would appear that a c impedance studies offer the greatest likelihood of resolving present uncertainties about the fundamental nature of mechanisms of corrosion rate control in chloride-bearing concretes and similar media Their major advantage in this respect lies in the ability to yield information concerning dynamic characteristics of the system, revealing in particular the influence of diffusion-controlled processes on the response of an embedded electrode In view of the importance which has been ascribed to diffusional effects with respect to the mechanism of chloride-induced pit growth, it would be of considerable interest to compare a c impedance data obtained in various concrete or hardened cement paste systems, for which chloride diffusion kinetics have been studied directly

Implications for the practical situation

We have, so far, concentrated on developments of a better basic understanding of the mechanisms of corrosion of steel in concrete However, there is also an important practical need for such better understanding The minority of reinforced concrete structures which deteriorate prematurely as a consequence of reinforcement corrosion do so as a result of poor constructional practices, inadequate design and, on occasions, poor quality of materials While codes of practice lay considerable emphasis on design and general aspects of quality control, in respect to the use of marginal quality concrete materials they give little detailed guidance, it is here therefore where a better knowledge of corrosion protection mechanisms especially in relation to the microstructure of the matrix, interactions of contaminants (particularly chloride and sulphate) and the role of water/cement ratio and curing could well lead to wider usage of poorer quality indigenous materials without adverse effects on durability At the same time, building owners will continue to repair existing deteriorated reinforced concrete structures and better understanding of the performance of repairs is urgently needed[56] This is particularly the case in relation to three factors (1) The diffusion of chloride through concrete (both in marine structures and where contact between chloride contaminated and uncontaminated units occurs) (2) The redistribution of anodic and cathodic sites in the concrete following repair (particularly in the substrate concrete) (3) The electrochemical interaction between repair and substrate materials and its influence on corrosion at the interface

Conclusions

Work should be undertaken in the following areas to improve practical understanding of mechanisms of corrosion and protection of steel in concrete

- Continued examination of the pore liquid in concrete at various stages of hydration and in different environmental conditions of curing and exposure
- Further investigations of the distribution of chloride (at a range of concentrations) between the solid and solution phases of hydrating cements and the relationship of this distribution to cement type
- Fundamental electrochemical studies of the performance of steel electrodes in electrolytes simulating pore solutions and complementary electrochemical studies on steel/concrete systems
- Studies of the diffusion rates of chloride through uncontaminated cement pastes and through those contaminated with chloride and/or sulphate
- Studies of oxygen diffusion through cement pastes and concretes, particularly in underwater exposure
- Development of improved electrochemical monitoring techniques for laboratory experimentation and *in situ* measurement of corrosion rates in structures, which accommodate difficulties arising from the high electrolytic resistance of concrete, and from crevice effects at electrode edges
- Investigation of the mechanisms of macrocell development in chloride contaminated concrete, especially in relation to concrete repairs

Improved understanding in these areas would have wide scope for application to the construction of durable reinforced concrete structures in adverse conditions and to the evolution of successful repair techniques The economic benefits to be gained from solution of the problems outlined above are therefore substantial

This review is based in part upon research carried out at the Building Research Establishment of the Department of the Environment and this paper is published by permission of the Director C L P gratefully acknowledges the financial support received from the SERC

1 *The Durability of Reinforced Concrete Buildings* Na ional Building Studies Spec Rep No 25 (HMSO, London, 1956)
2 Midgley, H C, Figg, J W & McLean, M J *Concrete* **7**, 24–26 (1973)
3 Brown, B L, Harrop D & Treadaway, K W J *Building Research Establishment Current Paper* CP 45/78 (BRE Garston, 1978)
4 Lea, F M *The Chemistry of Cement and Concrete* 3rd edn, 71 (Arnold, London, 1970)
5 Diamond, S *Proc Conf on Hydraulic Cement Pastes Their Structure and Properties*, Sheffield, 2–30 (Cement and Concrete Association, Slough, 1976)
6 Roberts, M H *Building Research Establishment Current Paper* CP 61/68 (BRE, Garston, 1968)
7 Peguin, P, Rubaud, M, Longuet, P & Zelwer A *Cah Cent scient tech Batim* No 130, Cahier 1109 (1972)
8 Longuet P, Burglen, L & Zelwer, A *Revue Matér Constr Trav publ Cim Betons*, **676**, 35–41 (1973)
9 Barneyback, R S Jr & Diamond, S *Cem Concr Res* **11**, 279–285 (1981)
10 Pourbaix, M *Atlas d'Equil bres Electrochimiques* 307 (Gauthier-Villars, Paris, 1963)
11 Hausmann D A *Mater Prot* **6**, 19–23 (1967)
12 RILEM Technical Committee CRC Report *Mater Struct Test Res* **9**(51), 187 (1976)
13 Gjorv, O E, Vennesland, O & El Busaidy, A H S in *Proc Corrosion/76*, Pap No 7 NACE, Houston, 1976)
14 Roberts, M H *Building Research Establishment Information Paper* IP 6/81 (BRE, Garston, 1981)
15 Treadaway K W J *Proc Symp on Corrosion of Steel Reinforcements in Concrete Construction* London, 1–14 (Society of Chemical Industry, London, 1979)
16 Page, C L *Bull Inst Corros Sci Technol* **77**, 2–7 (1979)
17 British Standards Institution Code of Practice CP 110 1972 Part 1, *The Structural Use of Concrete* (as amended May 1977)
18 Fookes, P G & Collis, L *Concrete* **10**, 14–19 (1976)
19 Crookes R N *Proc Symp Corrosion of Steel Reinforcements in Concrete Construction*, London 31–42 (Society of Chemical Industry, London, 1979)
20 Gjorv, O E & Vennesland, O *Cem Concr Res* **9**, 229–238 (1979)
21 Browne, R D & Geoghegan M P *Proc Symp Corrosion of Steel Reinforcements in Concrete Construction* London 79–104 (Society of Chemical Industry, London 1979)
22 Collepardi, M Marcialis, A & Turriziani, R *Il Cemento* **69**, 143–150 (1972)
23 Ost, B & Monfore, G E *J Res Dev Labs Portld Cem Ass* **8**, 46–52 (1966)
24 Cottringer, P & Kendall, N C *Concrete* **22**, 150–155 (1923)
25 Everett, L H & Treadaway, K W J *Building Research Establishment Information Paper* IP 12/80 (BRE Garston, 1980)
26 Gouda, V K *Br Corros J* **5**, 198–203 (1970)
27 Page, C L *Nature* **258**, 514–515 (1975)
28 Lea, F M & Watkins, C M *National Building Studies Research Paper* No 30, (HMSO, London, 1960)
29 Roberts, M H *Mag Concr Res* **14**, 143–154 (1962)
30 Richartz, W *Zem -Kalk-Gips* **22**, 447–450 (1969)
31 Verbeck, G J *Am Concr Inst Publ* SP 49, 21–38 (1975)
32 Building Research Establishment *The Structural Condition of Intergrid Buildings of Prestressed Concrete* (HMSO London, 1978)
33 Vassie, P R W *Bull Inst Corros Sci Technol* **81**, 11 (1980)
34 Henriksen, J F *Corros Sci* **20**, 1241–1249 (1980)
35 Gronvold, F O, Preece, C M & Arup, H *Proc 8th Int Congr Metallic Corrosion*, Mainz, **2**, 1800–1804 (1981), *Steel in Concrete Newsletter* No 3, 3–4 (1979)
36 Wilkins, N J M & Lawrence, P F *Proc Symp Corrosion of Steel Reinforcements in Concrete Construction*, London, 105–118 (Society of Chemical Industry, London, 1979)
37 Wilkins, N J M *Bull Inst Corros Sci Technol* **81**, 8 (1980)
38 Gronvold, F O *Steel in Concrete Newslett* No 4, 4–5, (1979)
39 Collepardi, M, Marcialis, A & Turriziani R *J Am ceram Soc* **55**, 534–535 (1972)
40 Page, C L, Short, N R & El Tarras, A *Cem Concr Res* **11**, 395–405 (1981)
41 Short, N R & Page, C L *Proc 8th Int Congr Metallic Corrosion* Mainz, **2**, 1767–1772 (1981)
42 Arup, H *Steel in Concrete Newslett* No 7 3–4 (1981)
43 Mitheux, I P thesis, New South Wales Institute of Technology (1976)
44 Page, C L *Paper presented at int Seminar on Electrochemistry and Corrosion of Steel in Concrete* Copenhagen (1979)
45 Baumel, A & Engel, H -J *Arch EisenhuttWes* **30**, 417–428 (1959)
46 Kaesche, H *Zem -Kalk-Gips*, **12**, 289–305 (1959)
47 Lewis, D A *Proc 1st Int Congr Metallic Corrosion*, London, 547–555 (1961)

48 Gouda, V K & Monfore, G E *J Res Dev Labs Portld Cem Ass* **7**, (3), 24–31 (1965)
49 Gouda, V K & Halaka, W Y *Br Corros J* **5**, 204–208 (1970)
50 Vassie, P R *Transport and Road Research Laboratory* Rep SR 397 (Crowthorne, TRRL, 1978)
51 Stratfull, R F *High Res Rec* **433**, 12–21 (1973)
52 Gouda V K, Shater, M A & Mikhail, R Sh *Cem Concr Res* **5**, 1–13 (1975)
53 Locke, C E & Siman, A *ASTM Publ* STP 713, 3–16 (1980)
54 Dawson, J L, Callow, L M, Hladky, K L & Richardson, J A in *Proc Corrosion/78* (Paper No 125 NACE, Houston, 1978)
55 King, R A, Dawson, J L & Geary, D *Proc Symp Corrosion of Steel Reinforcements in Concrete Construction*, London, 135–143 (Society of Chemical Industry, London, 1979)
56 O'Brien, T *Proc Instn civ Engrs*, Part 1, **68**, 399–408 (1980)

ARTICLES

Identification of resonance features within the rings of Saturn

J. B. Holberg*, W. T. Forrester* & Jack J. Lissauer†

* Earth and Space Sciences Institute, University of Southern California, Tucson Laboratories, Tucson, Arizona 85713, USA
† University of California, Departments of Astronomy and Mathematics, Berkeley, California 94720, USA

The Voyager 2 UV spectrometer observed a stellar occultation by the rings of Saturn, which located ring features with an accuracy of 12 km A high-resolution (3 km) optical depth atlas of the rings shows at least nine features, including four density wave patterns, identified with satellite resonances Analysis of these density wave patterns yields the first surface mass densities for the A ring and, together with our optical depth atlas, a total ring mass of 6.4 × 10⁻⁸ Saturn masses

IT has long been customary to link certain gaps and boundaries observed in the rings of Saturn to the locations of resonances with several of Saturn's satellites Most notable in this regard has been the association of the outer edge of the B ring with the location of the Mimas 2 1 resonance The vastly increased spatial resolution of the rings available from Voyagers 1 and 2 has made it possible to pursue this activity with much more confidence, however, the new clearer picture of the rings is complicated by two factors First, the rings contain a vast number of radial features on all scales Second, the newly discovered inner satellites have greatly increased the number of potential resonances with which one has to contend Both of these factors have led to the suggestion of numerous possible associations[1–4] To resolve any ambiguities a further increase in spatial resolution, together with a reliable distance scale for the rings is necessary Both of these criteria are met by the stellar occultation experiment performed by Voyager 2

Some of the first Voyager 1 images of the rings were used by Collins *et al*[1] to produce a list of observed features and possible corresponding resonance locations involving Titan, Iapetus, Hyperion, Mimas and the co-orbital satellites 1980S1 and 1980S3 The uncertainty in the absolute location of ring features, however, was ±500 to ±1,000 km Later Voyager 1 and 2 images showed wave-like patterns associated with several A ring features These features were suggested as being near to areas where two or more resonances coincided[2] A more positive identification was provided by Cuzzi *et al*[5] who analysed Voyager 1 images of the Cassini division showing density waves due to the Iapetus 1 0 resonance Further evidence of density waves associated with the 1980S1 2 1 resonance was provided by the Voyager photopolarimeter experiment[6] which obtained stellar occultation data simultaneously with the UV spectrometer

The ring features which we describe here are located near the inner Lindblad resonances with the external satellites of Saturn Many theoretical consequences of such resonances, including tightly wound spiral density waves propagating radially outward from the resonance location, were first discussed by Goldreich and Tremaine[7] (also see refs 5, 8)

In general these resonances are described by a ratio of two integers of the form $l/(m-1)$, which represents the approximate ratio of the mean orbital motion of a satellite to that of a ring particle Resonances of importance in Saturn's rings can be classified into two types[1] Type I resonances, where $l=m$, are usually the strongest and are independent of satellite orbital eccentricity, while type II resonances, where $l=m+1$, are weaker, being proportional to the first power of satellite orbital eccentricity A special type I resonance case occurs when $l=m=1$ Here the rate of the apsidal advance of a ring particle's orbit equals the mean orbital motion of the satellite We will refer to the three resonances of this variety present in the rings as apsidal resonances An additional class of resonance considered here involves resonances in the rings due to the rotation of Saturn[4] Such resonances depend on an azimuthal asymmetry in the gravitational figure of Saturn

The locations of the resonances used here are those of Lissauer and Cuzzi[8] who have computed the locations and strengths of expected satellite resonances using the zonal harmonics of Saturn's gravitational potential through J_6 These locations differ slightly from those given by Collins *et al*[1] and except for the apsidal resonances are believed to be accurate to within ±10 km Expected resonance locations are shown in Fig 1 together with normal optical depths (τ) of the A, C and portions of the B ring observed by the UV spectrometer

δ Sco occultation

During the Voyager 2 Saturn encounter the UV spectrometer[9] observed an occultation of the bright B0 5 IV star δ Scorpii by the rings of Saturn The occultation began as the star emerged from the atmosphere of Saturn at 23 45 07 UTC 25 August 1981 and continued for 130 min During this time the apparent path of the star traversed the entire ring system from the D ring to beyond the F ring with a mean radial velocity of 10 3 km s^{-1} Throughout the occultation the UV spectrometer obtained a complete 912–1,700 Å spectrum of δ Sco every 0 32 s, yielding a resolution of ~3 3 km in the rings A preliminary discussion of the UV spectrometer ring occultation results is given elsewhere[10] The data from which the normal optical depths of the rings were derived are the result of summing an entire UV spectrum to yield a single intensity measurement at an effective wavelength of 1,125 Å These intensity data were then corrected for instrumental background and the effects of spacecraft limit cycle motion Normal optical depths have been computed from the ratio of the attenuated to the unattenuated signal and corrected for the 28 71° angle between the line of sight to δ Sco and the plane of the rings Details of the data reduction will be presented elsewhere

0028–0836/82/190115—06$01 00

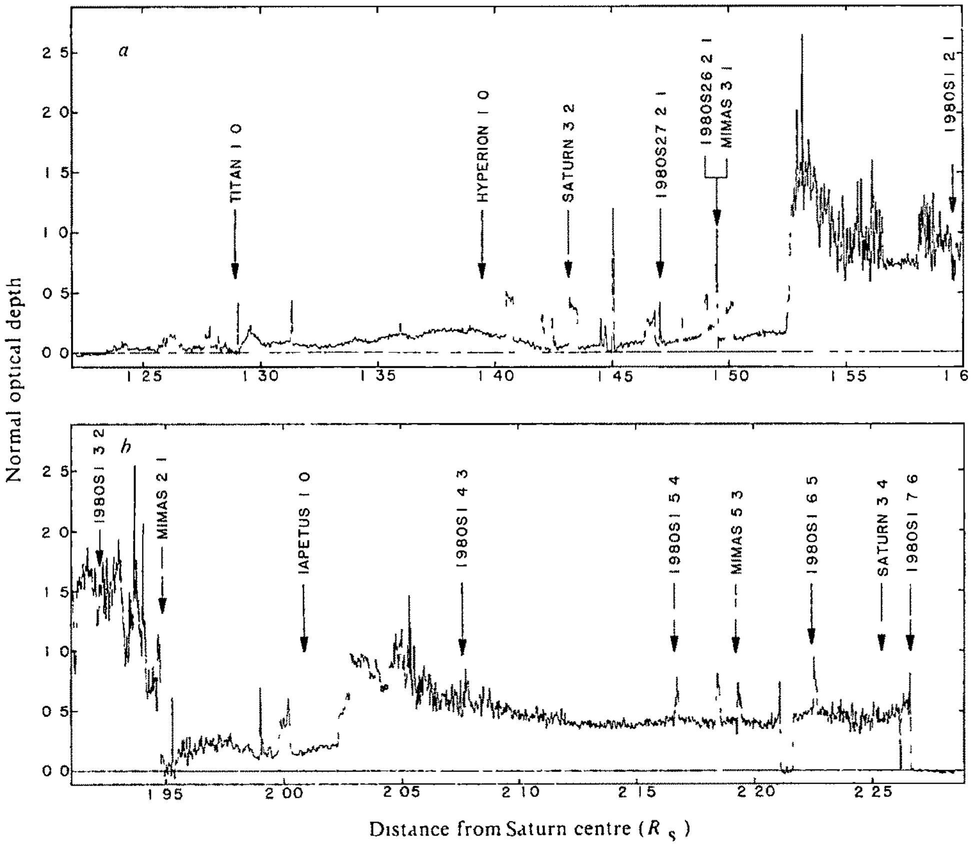


Fig 1 The normal optical depths in the rings observed by the Voyager 2 UV spectrometer These data, which have been plotted at a resolution of ~15 km, span the entire C and A rings and include those portions of the B ring containing resonances discussed in this report The expected locations of these resonances[4 8] are indicated by arrows *a*, The C ring and the inner portion of the B ring *b*, The outer portions of the B ring, Cassini division and the A ring Optical depths between 1 95 and 1 983 R_S are preliminary due to present uncertainty in the spacecraft limit cycle in this region The resulting uncertainty in the Cassini division mass (Table 2) is perhaps as large as 50% Precise ring boundaries observed by the UV spectrometer are given in Table 2

Many of the identifications of ring resonances presented here rest on the absolute radial locations obtained for the associated features A brief discussion of the accuracy of these determinations is therefore appropriate For a star occulted by the rings, the location of observed ring features depends only on the apparent position of the star, the precise trajectory of the spacecraft relative to Saturn coupled with the timing of events, and the location of Saturn's rotational pole The uncertainty in the apparent position of δ Sco, after corrections for proper motion, parallax and aberration of starlight is negligible (<0 1 km in the ring plane) The trajectory uncertainty projected on to the ring plane is of the same order as the uncertainty due to the UV spectrometer time resolution, 2 4 and 3 2 km respectively The dominant uncertainty, at present, is due to the location of Saturn's rotational pole Errors in the location of ring features produced by this factor are difficult to estimate, but use of various pole locations derived from Voyager trajectory analysis as well as pre-Voyager determinations cause location changes of <20 km in the rings

Fortunately, it is possible to compare directly the distance scale used here with the locations of features seen in two additional stellar ring occultations observed with the UV spectrometer from Voyagers 1 and 2 These are a Voyager 1 observation of ι Herculis and the Voyager 2 entrance of δ Sco through the opposite side of the rings Both of these occultations extended out to 1 43 R_S ($1R_S = 60{,}330$ km) The locations of all recognizable features in these three occultations are in mutual agreement to within 12 km except for one ringlet known to be eccentric from imaging data If we make the reasonable assumption that this agreement is not just coincidence but rather a consequence of symmetry in this part of the C ring, then we can be confident of our absolute distances to this level Ultimately it should be possible to reduce the locational uncertainties to 6 km or better

Detected and undetected resonances

The recently discovered small inner satellites of Saturn[2], 1980S1, 1980S3, 1980S26, 1980S27 and 1980S28 all possess numerous resonances within the ring system Because of the small mass of several of these bodies only a fraction of these resonances are important Of these, the type I resonances due to the leading, more massive, co-orbital 1980S1 are trongly evident in the rings Figure 2 shows full resolution data for the six type I 1980S1 resonances seen in Fig 1 Three of these resonances display characteristic density wave structure (2 1, 5 4, 6 5) while the 3 2 and 4 3 resonances are more chaotic These will be discussed more fully together with other resonances in high optical depth regions of the rings, however, one point is unique to 1980S1 resonances The interchange of the orbits of 1980S1 and 1980S3 which occurs every 4 yr [11], changes the resonance locations of 1980S1 by 10–15 km A 'kink' in the wave train of approximately this magnitude will develop (S Tremaine, personal communication) and propagate outward at the wave's group velocity[7] In the ~960 days (ref 11) between the last interchange and our observations, these waves have travelled roughly 60 km from the 2 1 resonance of 1980S1 and over 100 km from the other type I resonances This effect may cause a slight error in our subsequent determination of the surface density at the 2 1 resonance The 7 6 resonance is coincident with the outer edge of the A ring The smaller co-orbital, 1980S3, has resonances which occur 30–45 km inward of the corresponding 1980S1 resonances These resonances are not apparent in our data

Features due to resonances with the two F ring shepherd satellites 1980S26 and 1980S27 are also present in our data (Fig 1) Many of the >50 type I shepherd satellite resonances in the rings are noticeable, particularly in the outer A ring We have included in this discussion only the 2 1 resonances with each of these satellites because they occur in low optical depth regions Because of their number and lower amplitude, discussions of the remaining 1980S26 and 1980S27 resonances will be presented elsewhere There are ~100 1980S28 type I resonances present in the rings, near the outer edge of the A ring they reach a separation of 9 km At present there is no evidence of any features due to resonances with this small satellite

The occultation data show clear evidence (Fig 3) of features corresponding to all three of the strong Mimas resonances present in the ring system The strongest and best known is the 2 1 resonance which has long been believed to be coincident

with the outer edge of the B ring The two additional Mimas resonances are the only known examples of type II resonances seen in the rings Apparently, the 0 02 eccentricity of Mimas is large enough to allow Mimas to exert sufficient torque to create observable features of the 5 3 and 3 1 resonance locations

The only apsidal resonance which we can definitely detect is the Titan 1 0 which occurs in the C ring (Fig 4) Hyperion and Iapetus, being much less massive and more distant than Titan, exert significantly less torque at their apsidal resonances so that our failure to observe features at these resonances is not surprising The density waves described by Cuzzi *et al*[5] near the Iapetus apsidal resonance are of too low an amplitude to be seen in our data The stated location of the wave source[5], however, is seen in Fig 1 as the inner edge of the band of low τ containing the Iapetus resonance The wave source and the resonance location are separated by ~370 km

Finally, we have examined our data for the presence of ring features associated with the predicted location of resonances with Saturn's rotational period The possibility of such resonances has been suggested by Franklin *et al*[4], who tentatively identified the locations of the 3 2 and the 3 4 rotational resonances with two prominent gaps in the rings seen in Voyager 1 images Our data (Fig 1) show the 3 2 resonance to be located at the outer edge, rather than the centre, of the gap suggested by Franklin *et al*[4] The 'gap' itself has a distinct non-zero optical depth and is similar in appearance to several other inter-ringlet regions in this part of the C ring The precise location of the 3 2 resonance falls at the inner edge of the 230 km-wide ringlet located at 1 434 R_S Doubt as to the significance of any link between the resonance and either the gap or ringlet is increased by the failure to observe the presence of the related 3 4 resonance Franklin *et al*[4] associated the 3 4 resonance with the prominent narrow gap between the Encke division and the outer edge of the A ring We find this gap to be located at 2 2627 R_S, 440 km outside the location of the resonance but within the ±700 km uncertainty of the Franklin *et al*[4] observational data At the location of the resonance itself we see no convincing evidence of any associated structure

Resonances in low τ regions

The morphology of ring features associated with the strongest resonances appears to be controlled by the local optical depth

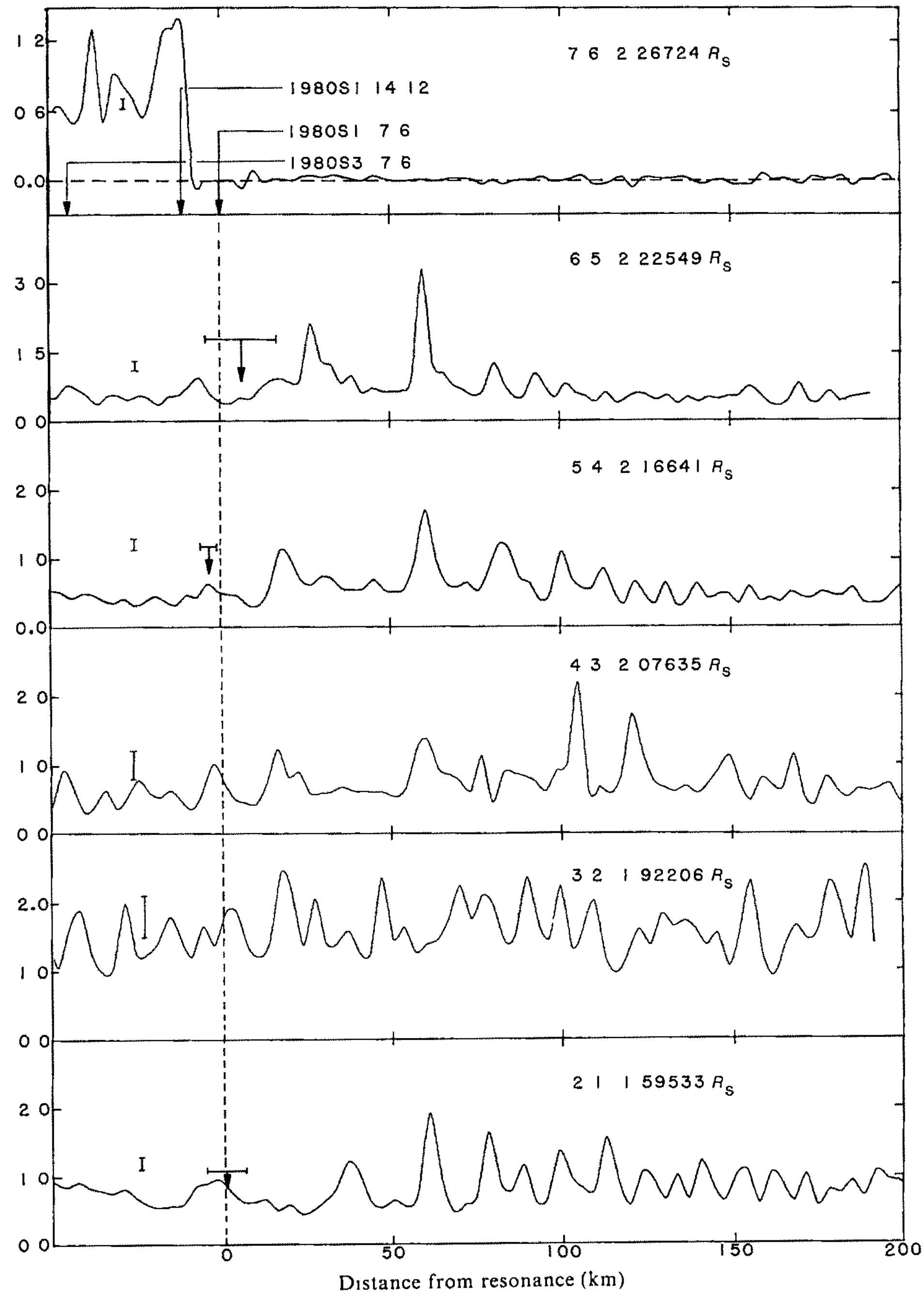


Fig. 2 The normal optical depths for regions located near the six type I resonances due to the co-orbital satellite 1980S1 The origin of the horizontal scale corresponds to the expected location of each resonance Spiral density wave patterns with characteristic damped waves of decreasing wavelength are seen exterior to the 2 1, 5 4 and 6 5 resonances The appearance of the rings at the 3 2 and 4 3 resonances is more chaotic Vertical error bars show the single sample noise level at the indicated τ Arrows and horizontal error bars show the resonance locations deduced from analysis of the density waves

Table 1 Ring features corresponding to resonance locations

Source	Resonance	Expected location (R_S)*	Surface density σ (g cm^{-2})	Opacity K (cm^2 g^{-1})	Type of feature
1980S27	2 1	1 47043			Narrow ringlet/gap
1980S26	2 1	1 49456			Narrow ringlet/gap
1980S1	2 1	1 59535	70±10	0 01±0 002	Density waves[6]
	3 2	1 92206			?
	4 3	2 07635			Disturbance
	5 4	2 16641	56†±12 (100)	0 008±0 002	Density waves
	6 5	2 22549	40†±6 (110)	0 01±0 002	Density waves
	7 6	2 26724			Outer edge of A ring
Mimas	3 1	1 49503			Narrow ringlet/gap
	2 1	1 94847			Outer edge of B ring (narrow ringlet/gap)
	5 3	2 19286	45†±8 (80)	0 01±0 002	Density waves
Titan	1 0	1 28951			Eccentric ringlet/gap
Hyperion	1 0	1 39499			None
Iapetus	1 0	2 00867			? (Density waves[5])
Saturn	3 2	1 43163			Outer edge of gap?
	3 4	2 25543			None

* $1R_S$ = 60,330 km

† Measured outside region of enhanced surface density, densities in parentheses refer to region of enhancement

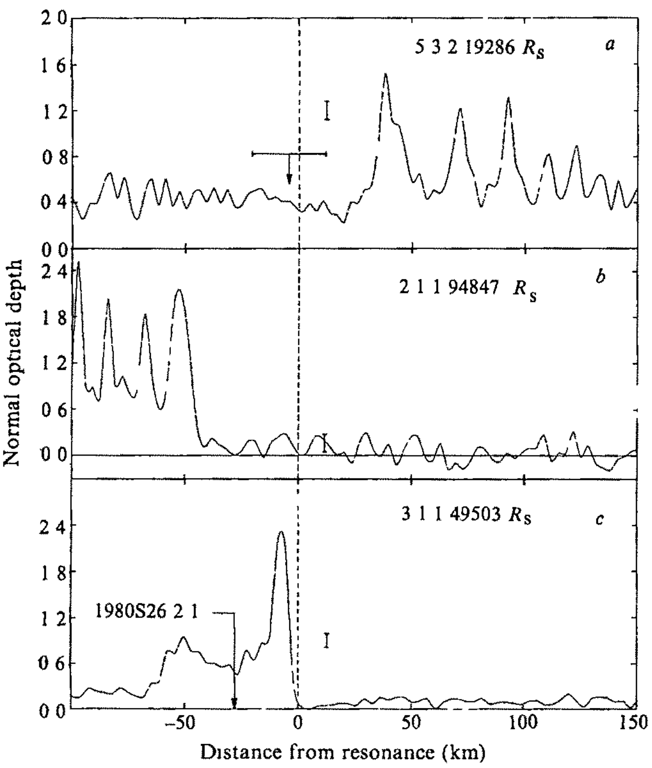


Fig 3 The normal optical depths for regions located near the three Mimas resonances The origin of the horizontal scale corresponds to the expected location of each resonance The 5 3 resonance (a) is an excellent example of density waves propagating outwards from a resonance location The 2 1 resonance (b) is located 48 km outside the outer boundary of the B ring but within the region over which this boundary is observed to vary[3] The 3 1 resonance (c) is located at the outer edge of a narrow ringlet in the C ring and may interact with the 1980S26 2 1 resonance to produce the sharp peak in the ringlet Note the near zero optical depth exterior to this ringlet Vertical error bars show the single sample noise level at the indicated τ The arrow and horizontal error bar near the 5 3 resonance indicate the location of the resonance deduced from analysis of the density waves

of the region in which the resonances reside Looking first at regions of low optical depth (τ<0 2), we note that there are three locations containing major resonances in the C ring (Fig 1) The Titan apsidal resonance lies farthest inward, while the 2 1 resonances with 1980S27 and 1980S26 are located near the outer edge of the C ring, the latter of these two being very nearly coincident with the Mimas 3 1 resonance (see Fig 3) All three of these resonances occur in regions characterized by a gap of lower than average optical depth containing a narrow optically thick ringlet The gap associated with the Titan apsidal resonance is several hundred kilometres wide and completely empty The 1980S27 resonance is situated in a region of lower than average, but apparently non-zero, optical depth while the region just exterior to the 1980S26 2 1 and Mimas 3 1 resonance is devoid of ring material to within our level of detection

The Titan apsidal resonance, the only one of these features for which there exist multiple UV spectrometer observations, is associated with a narrow, double-peaked, optically thick ringlet of substantial eccentricity[3,10] The two Voyager 2 UV spectrometer δ Sco occultations locate this ringlet at 111 km and 67 km exterior to the Titan apsidal resonance at longitudes separated by 149° and 3 53 h of time The 1980S27 2 1 resonance and the close pair formed by the 1980S26 and Mimas resonances are also associated with narrow, double-peaked, optically thick ringlets The Mimas 2 1 resonance at the inner edge of the Cassini division also manifests itself in a gap containing an opaque double-peaked ringlet In this case, not only is the ringlet eccentric but the inner edge of the gap also varies in radius and is shaped like an ellipse centred on Saturn[3] In the UV spectrometer data the outer edge of the B ring was located 48 km inside the theoretical location of the Mimas 2 1 resonance At the time of the stellar occultation Mimas, which controls the orientation of the B ring's outer edge, lagged the point of observation in the rings by a phase of 194° This datum, when combined with imaging measurements, should give a more precise picture of the location and shape of the outer edge of the B ring

We do not understand fully why resonances which occur in regions of low optical depth manifest themselves as gaps containing narrow double-peaked ringlets Resonantly induced collisions are more likely to clear gaps in very low τ regions[8], however, t is questionable whether τ is sufficiently low for this process to occur even in the C ring, also in most cases the size of the gap is larger than the collision predicted width Density

waves may also be more effective in clearing a gap in low surface density regions because they will quickly become nonlinear, depositing their negative angular momentum over a smaller region[7] A complete theory of these gaps must involve the opaque ringlets embedded within them These ringlets, which seem to be characterized by a double-peaked structure similar to some of the uranian rings[12], also occur within some non-resonant gaps in the C ring Imaging data will be important for the determination of the azimuthal structure of these ringlets as well as the gap edges

Resonances in regions of medium and high τ

The predominant effect of resonances in regions where τ is between 0 4 and 1 0 is the excitation of density waves We have selected for analysis the four clearest examples of such density waves, those associated with the Mimas 5 3 resonance (Fig 3), and the 6 5, 5 4, and 2 1 resonances with 1980S1 (Fig 2), the 2 1 having been previously studied by Lane *et al*[6] The two other 1980S1 type I resonances are noisier in our data The 4 3 occurs in the inner and most opaque part of the A ring and clearly causes some stirring but a clean wave pattern is not present The 3 2 occurs in a high τ region of the B ring and is not immediately evident in our data At present it is unclear whether the lack of observable density waves in the highest τ regions is due to their absence, their confusion with the crowded ringlet structure of the B ring[2] or simply our low signal level when τ becomes very large

One of the theoretical predictions concerning the development of density waves near a resonance is the opening of a gap just exterior to the resonance[7] Not only do we find no evidence of density waves clearing gaps, the average τ in resonance regions is in general higher than that of surrounding regions Density waves seem to damp and thereby deposit negative angular momentum over a region of several hundred kilometres, so it may be that material is brought inwards towards a resonance faster than it can be cleared from the region by damping of the first wave An equilibrium may be reached when surface mass density (σ) is increased enough for viscous diffusion out of the region to create a balance Another possibility is that collisions alter the particle size distribution by increasing the numbers of small particles, thereby increasing τ

The density waves excited at the Mimas 5 3 and at the 1980S1 type I resonances are also quite highly nonlinear (shocked) for the first few wavelengths These waves have a distinctly non-sinusoidal character with a sharp peak at the location of the shock and a broad flat trough (see ref 13 for a theoretical analysis of nonlinear density waves in galaxies)

Density waves are a very useful tool for deducing such ring properties as surface mass density and viscosity, unmeasurable by any other currently available means[5,6], however, some caution must be exercised First, the shocking of the waves removes a great deal of their angular momentum, thus viscous damping plays a minor part and any measure of viscosity from wave damping would only constitute a crude upper bound (a slight phase shift also results when the waves shock, thus changing the apparent wavelength between crests, however, this small effect should not significantly affect our estimates of σ derived in equation (1)) Second, the mean (wavelength-averaged) surface density can be non-uniform over the region of a particular wave train (we see evidence of this in the A ring) This may lead to problems in the determination of the resonance location as the mean resonance wavelength decreases as σ/d, where d

Table 2 Ring mass

Ring	Boundaries (R_S)	Mass ($10^{-8} M_S$)
C	1 2342–1 5259	0 2
B	1 5259–1 9477	5 0
Cassini division	1 9477–2 0232	0 1
A	2 0232–2 2676	1 1
	Total	6 4

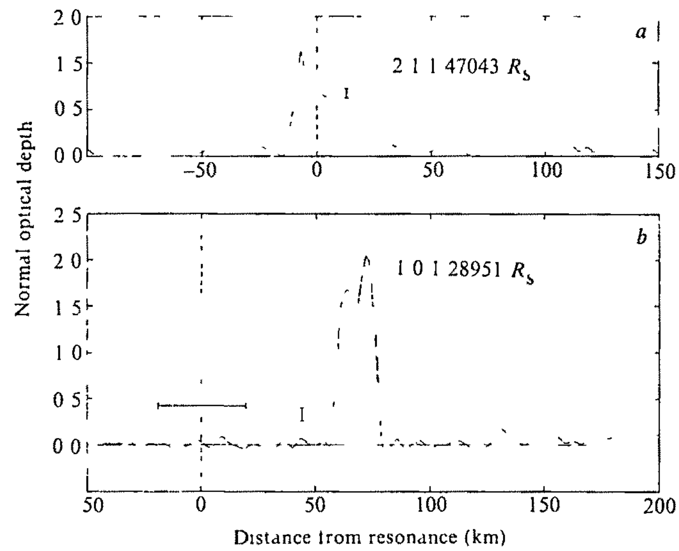


Fig 4 The normal optical depths for regions located near the 1980S27 2 1 (a) and Titan apsidal resonances (b) The origins of the horizontal scales correspond to the expected location of each resonance The horizontal error bars corresponding to the Titan apsidal resonance location are due to the uncertainty in Saturn's zonal harmonics[15] The vertical error bars in each panel show the single sample noise level at the indicated τ

is the distance from the resonance[14] A more severe problem would develop from changing σ if we were to try to determine viscosity, because the wave amplitude can change significantly with changes in σ

With these cautions in mind we proceed to determine σ and the opacity K via the following relations

$$\sigma(\mathrm{g\,cm^{-2}}) = \frac{0\ 325(m-1)\lambda d}{R^4} \quad (1)$$

$$K(\mathrm{cm^2\,g^{-1}}) = \frac{\tau}{\sigma} = \frac{R^4}{0\ 325(m-1)d}\frac{\tau}{\lambda} \quad (2)$$

where R is the resonance location in Saturn radii, and the wavelength λ and the distance from the wave source d are measured in kilometres We find that all three of the A ring density waves exhibit strong enhancements of local (wavelength-averaged) τ over the initial 100 km of each wave train, indicating that density waves cause changes in local σ or K We therefore analysed our data separately assuming constant α and constant K The assumption of constant σ (equation (1)) does not provide an adequate fit to the observed wavelengths as a function of d The assumption of constant K does, however, give a satisfactory fit to the data On this assumption, the quantity τ/λ in equation (2) is determined from the observed wavelength-averaged τ for each wave The A ring results for K and σ given in Table 1 and the wave source locations shown in Figs 2 and 3 are based on the constant opacity assumption

Ring boundaries and ring mass

As we have noted previously, the 7 6 resonance is in extreme proximity to the outer edge of the A ring (Fig 2) We feel it is not chance that the second strongest resonance in the rings should coincide with the outer edge of the ring system while the strongest resonance, the Mimas 2 1, lies at the outer boundary of the B ring Satellites exert a torque on particles at resonances, removing angular momentum from them and thus pushing the particles closer to Saturn A similar satisfactory explanation of the inner boundaries of the major rings is not yet available

A detailed study of the outer edge of the A ring requires azimuthal coverage available only from imaging data We would expect a seven-lobed pattern in analogy with the two-lobed

oval pattern seen at the Mimas 2 1 resonance Observation of more frequently spaced undulations in the outer edge of the A ring boundary would be indicative of the effects of the higher m-order resonances present in the region (Fig 2)

Our determinations of surface mass density at three points in the A ring constitute the first such measurements for this ring Our single mass density measurement in the B ring at the 1980S1 2 1 resonance is in good agreement with the 60 g cm^{-2} at this resonance obtained earlier by Lane *et al*[6] All of our surface densities are consistent with a constant ring opacity of 0 01 $cm^2\ g^{-1}$, as are prevous measurements[5,6] Using this result, together with our optical depth atlas, we can estimate ring masses as follows

$$M_{\rm ring} = 2\pi \int_{R_i}^{R_0} \sigma r\, dr = 2\pi \int_{R_i}^{R_0} \frac{\tau r dr}{K} = \frac{2\pi}{K} \int_{R_i}^{R_0} \tau r\, dr \qquad (3)$$

where $M_{\rm ring}$ is the total mass of the ring, r is a radius in the ring and R_i and R_0 are the inner and outer ring boundaries, respectively The results of these integrations for the major rings are given in Table 2, together with UV spectrometer measurements of the ring boundaries Our determination of the B ring mass may be underestimated due to the fact that we have assumed a maximum optical depth of $\tau = 2\ 85$ over the 15% of the B ring where we see no stellar signal Measurements from other Voyager instruments may help to establish accurate optical depths in these regions The A ring may also be more massive if the opacities we determine at the resonances are higher than the rest of the A ring due to an enhancement of small particles produced by collisions in the shocked region of the density waves

We find the total mass of the ring system to be 6 $4\times10^{-8}\ M_S$, approximately the same mass as Mimas Assuming uniform ring opacity, this total mass could be in error by perhaps 30% The best previous estimate of the mass of the rings was the upper limit of $170\times10^{-8}\ M_S$ from the Pioneer 11 gravity results[15] Our much smaller measurement indicates that an actual measurement of ring mass is probably unobtainable from existing spacecraft gravity measurements

Conclusions

We have shown that resonances are associated with a wide range of morphological features in the rings In low optical depth regions such as the C ring and inner Cassini division they are linked with gaps containing narrow double-peaked ringlets, while in higher τ regions they appear as density waves and wave-like disturbances The two strongest resonances are coincident with the outer edge of the B ring and the outer edge of the A ring Known resonances have a major role in determining the structure of the A ring but can explain only a small fraction of the observed features in the B and C rings, where resonances are less frequent We have analysed the four density wave trains to determine their point of origin and the local surface mass density These estimates of surface mass density in combination with our optical depth atlas make possible an estimate of the total mass and mass budget of the ring system

We thank J N Cuzzi and F H Shu for helpful discussions This work was supported by the JPL, California Institute of Technology under NASA contract NAS 7-100 Additional support was provided by the Planetary Sciences Discipline of NASA's Office of Space Sciences grant NAGW-62 J J L was supported by NASA training grant NGT-05-003-800

Received 5 March, accepted 29 March 1982

1 Collins, S A *et al Nature* **288,** 439 (1980)
2 Smith, B A *et al Science* **212,** 163 (1981)
3 Smith, B A *et al Science* **215,** 504 (1982)
4 Franklin, F A, Colombo, G & Cook, A F *Nature* **295,** 128 (1982)
5 Cuzzi, J N, Lissauer, J J & Shu, F H *Nature* **292,** 703 (1981)
6 Lane, A L *et al Science* **215,** 537 (1982)
7 Goldreich, P & Tremaine, S *Icarus* **34,** 240 (1978)
8 Lissauer, J J & Cuzzi, J N *Astr J* (in the press)
9 Broadfoot, A L *et al Space Sci Rev* **21,** 183 (1977)
10 Sandel, B R *et al Science* **215,** 548 (1982)
11 Dermott, S F & Murray, C D *Icarus* **48,** 12 (1981)
12 Elliot, J L *et al Astr J* **86,** 444 (1981)
13 Shu, F H, Milione, V & Roberts, W W *Astrophys J* **183,** 819 (1973)
14 Shu, F H *Astrophys J* **160,** 99 (1970)
15 Null, G W *et al Astr J* **86,** 456 (1981)

Stratospheric condensation nuclei variations may relate to solar activity

D. J. Hofmann & J. M. Rosen

Department of Physics and Astronomy, University of Wyoming, Laramie, Wyoming 82071, USA

Observations of increases of stratospheric condensation nuclei suggest a photo-initiated sulphuric acid vapour formation process in spring in polar regions It is proposed that the sulphuric acid rapidly forms condensation nuclei through attachment to negatively charged multi-ion complexes and that the process may be modulated through variations in solar activity

THE possible connections between solar variability and the weather on the one hand and climate on the other have been widely debated[1,2] In studies of the correlation of solar activity with both weather and climate, there has been little or no experimental evidence to show by which mechanism minute changes in solar energy incident at the Earth's orbit may be transformed to observed weather or climate variations A direct coupling through variations in the solar constant has generally been discarded in favour of a triggering mechanism which utilizes charged solar corpuscular radiation The latter may take the form of solar control of galactic cosmic radiation through the waxing and waning of solar magnetic irregularities in the interplanetary medium The triggering mechanism may also appear as direct magnetospheric intrusion of solar flare accelerated charged particles and/or energetic electrons released from the magnetosphere by enhanced geomagnetic activity due to active Sun processes

The intermediary in the solar variation–weather connection may be the stratospheric ozone layer which can be depleted by the production of nitric oxide during charged particle bombardment (solar or galactic) of the upper atmosphere and thereby may affect the intensity of solar UV radiation in the Earth–atmosphere system It may also take the form of the creation of atmospheric ionized species through charged particle bombardment which alter the electrical characteristics of the atmosphere and thereby modulate the occurrence of thunderstorms

One of the more effective ways of altering climate, variations in cloudiness or the particulate content of the atmosphere, has perhaps the least amount of observational support as far as solar activity correlations are concerned Increased strato-

0028-0836/82/190120—05$01 00

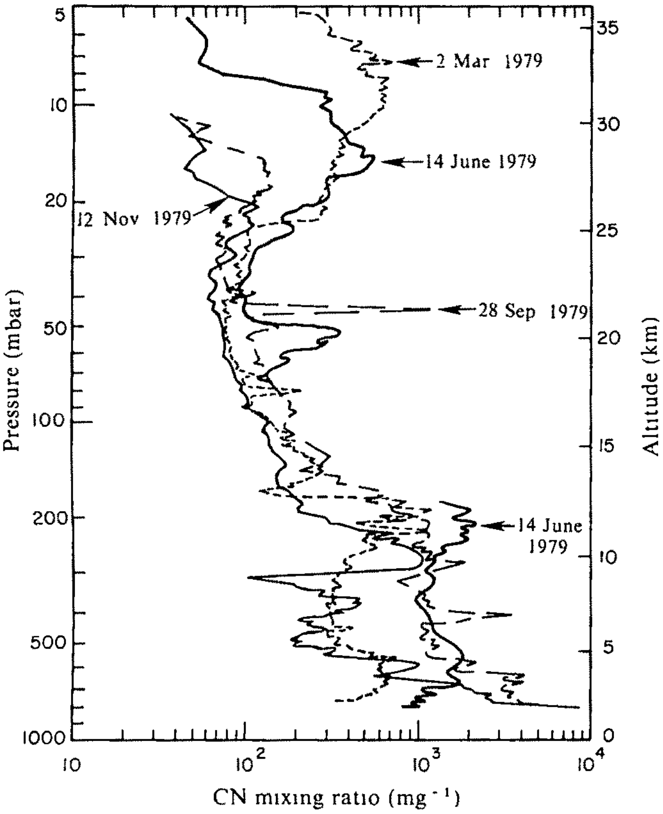


Fig. 1 Condensation nuclei ($r \geq 0.01$ µm) mixing ratio profiles during 1979 showing the time variation of the high altitude event

spheric particulates, due to volcanic eruptions, are capable of heating the stratosphere and cooling the surface Thus, any mechanism which can create material particles in the atmosphere is a likely candidate for climatic studies

Ney[3] suggested that the logical place to look for solar–weather effects was the solar modulation of galactic cosmic radiation Later Roberts and Olson[4] proposed that atmospheric ions may serve as condensation nuclei (CN) for the formation of cirrus clouds in the 10–12 km region, thus forming a solar activity connection with climate Dickinson[5] proposed that ions formed by the galactic cosmic radiation may be important in the formation of stratospheric sulphuric acid aerosol which, in turn, could nucleate cloud droplets in the upper tropospheric environment With primary ionization control ascribed to the galactic cosmic radiation, minimal ion density at solar maximum would suggest fewer growth sites, less competition for the available sulphuric acid vapour and thus larger though fewer particles at this time

The formation of stratospheric aerosol through ions has largely been neglected Thus, the stratospheric aerosol formation models of Rosen *et al*[6] and Turco *et al*[7] included only a tropospheric source of pre-existing CN which diffused into the stratosphere and served as sites for larger aerosol particle growth by accretion of sulphuric acid vapour This assumption of a tropospheric source of CN was justified by the measurements[8,9] wherein the CN mixing ratio peaked just below the tropopause and fell off sharply in the stratosphere However, the observation of an essentially constant, albeit low, mixing ratio from 20 km to the highest altitude of observation (~32 km) left some room for speculation that a high altitude source (for example, meteoritic) may be operative

The possibility that this source could be ionic in nature was not considered because even at typical stratospheric ion concentrations of 5,000 cm^{-3}, ion-induced nucleation cannot compete with heterogeneous nucleation on pre-existing particles for typical stratospheric particle concentrations[10] and only following major solar flare particle events (which are very rare), can ion induced nucleation be important

Recent observations have cast new light on this problem Following the initial stratospheric ion-mass spectrometer measurements[11], Ferguson[12] interpreted the results as indicating the clustering of ions, forming stable ion pairs, and suggested their possible importance in nucleation processes in the stratospheric sulphate layer

Arnold and Henschen[13] followed the initial positive ion measurements with measurements of negative ion clusters which suggested the presence of sulphuric acid in these clusters Arnold[14] suggested that attachment of free ions to these clusters could lead to larger multi-ion complexes (MIC), or that they may even mutually coalesce He suggested that formation of MIC and subsequent growth could result in a gas-to-particle conversion mechanism which does not require high gas supersaturations In the presence of a supersaturated gas, such as sulphuric acid vapour, MIC which have grown to embryonic size could act as condensation nuclei Arnold and Henschen[15] found a pronounced layer of MIC at 30 km having masses in excess of 327 AMU and labelled them 'pre-condensation nuclei' These new observations have prompted us to examine some of our recent stratospheric particle observations, the results indicating that a high altitude ion–CN relation may exist

Observations

We have recently reported[16] the observation of an unusual increase in the CN ($r \geq 0.01$ µm) concentration above 25 km which lasted about 6 months in 1979 (March through September) and some evidence for its repetition in March, 1980 at Laramie (41°N) Measurements were made using a balloon-borne thermal growth chamber, attached to a light scattering particle counter In 1979 the increase was peaked at ~32 km with a mixing ratio of ~650 mg^{-1} above a normal background of ~30 mg^{-1} at this altitude (a concentration of ~9 cm^{-3} above a background of ~0 4 cm^{-3})

No attempt was made to explain the nature of this event as it was not conclusive if the feature was really annual on the basis of the data at hand Earlier measurements probably would not have detected such an annual event, if it existed, because the CN growth chamber only functioned satisfactorily to an

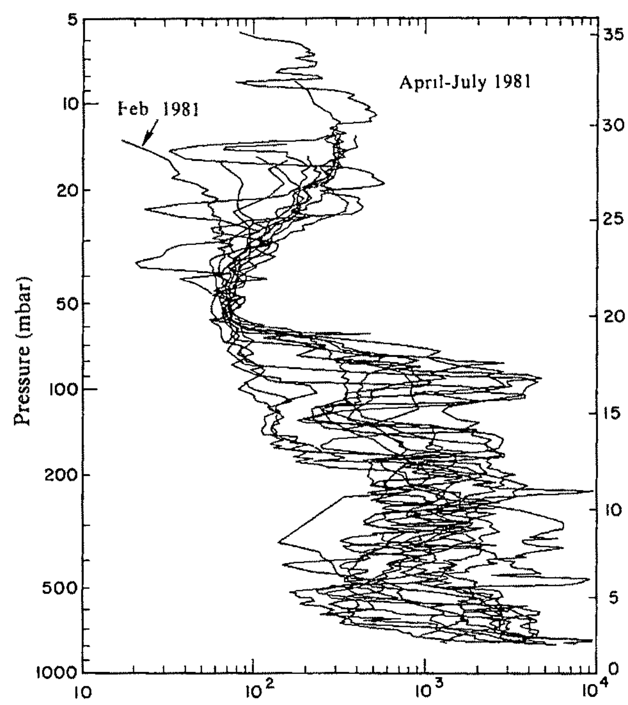


Fig. 2 Condensation nuclei ($r \geq 0.01$ µm) mixing ratio profiles during 1981 showing the high altitude event Enhanced CN layers in the 13–18 km region were due to the volcanic eruption of Alaid on 28 April 1981

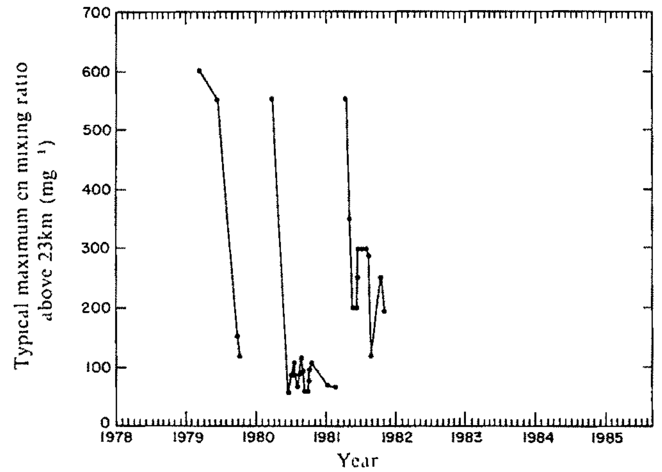


Fig. 3 Condensation nuclei ($r \geq 0.01$ µm) time variation showing the occurrence of the springtime events at Laramie.

altitude of 25 km (refs 8, 9). The 1980 investigations were complicated by the eruption of Mt St Helens in May 1980 and while numerous measurements were made, smaller balloons were used to study the 20–25 km effects of Mt St Helens and the 30–35 km region was not effectively studied until September when CN levels appeared normal[17].

Five soundings during 1979, shown in Fig. 1, indicated that the excess CN decayed exponentially with a e^{-1} time of 130 days. As the layer decayed, the mixing ratio peak appeared at successively lower altitudes, finally blending into the background at about 25 km. Initial attempts to explain the event as being due to the burnup of a large piece of space junk in the atmosphere were rapidly discarded. One tonne of such material, distributed uniformly in a 10-km thick cloud with a particle concentration of 10 cm^{-3} (mixing ratio of 600 mg^{-1} at 30 km) would have a horizontal extent of only 38 km for particle sizes of 0.07 µm (typical CN sizes in the unperturbed stratosphere) and 700 km for sizes of 0.01 µm, a lower limit of detection for the instrument. Such a small cloud would not have been detected on all five CN flights conducted in 1979. Considerations of a large meteor would result in the same conclusions. The observation that the CN mixing ratio displays a well defined maximum between 25 and 35 km further rules out objects entering the atmosphere from space. The latter, if observable, would display a broad mixing ratio maximum extending to great altitudes. It was considered more likely that the layers observed in 1979 were all derived through *in situ* production at their origin but that variable transport from a source region resulted in the observed characteristics.

In 1981 the high altitude enhancement did not seem to be present in February, at least in the 25 km region, but it was very apparent in April and persisted through 12 subsequent soundings. The data are shown in Fig. 2. Some complications resulted from the volcanic eruption of Alaid in the Kurile Islands in April 1981 but these disturbances were restricted to the 13–18 km region[18]. Figure 3 shows the high altitude CN mixing ratio versus time and clearly demonstrates the periodic nature of the events.

Discussion

The observation that the CN events are seasonal, favours variable transport to midlatitudes from a remote source region as there is no known seasonally varying production source at midlatitude. Little is known about the nature of meridional transport or its seasonal variations at altitudes in the 30 km region. If one associated possible meridional transport with regions of high wind speed, then likely candidates at 30 km in the Northern Hemisphere are the polar vortex region at 50–80° N during December–February ($v = 20$–40 m s^{-1}) and the region 0–30° N during June–August ($v = 20$ m s^{-1}) (ref. 19). In view of our observation that the enhanced CN layers appear in March, the polar regions appear to be the most promising. In addition, soundings in Panama (9° N) in 1976 and 1977 and in Brazil (5° S) in 1978 did not reveal excess CN mixing ratios in the upper stratosphere although the CN detector used then was suspect above 25 km (ref. 9).

In view of the more probable source region for the CN enhancements being at high latitude, and the fact that multi-ion complexes have been observed in a layer at the same altitude as the CN enhancements appear, we feel that there may be a logical connection between the two and thus possibly a connection between solar activity and stratospheric aerosol formation.

The enhancements are not observed until March at the earliest while the polar vortex 30 km wind system is operative throughout winter. This suggests that excess CN may not be formed at 30 km during the polar winter night but begin to form in spring. Thus a sunlight effect in the formation of CN appears to be a possible mechanism.

As large aggregate negative ion clusters contain sulphuric acid, and since stratospheric CN is largely volatile at 150 °C, suggesting a sulphuric acid composition[16], it seems possible that a sunlight effect may be manifested through the photochemical production of sulphuric acid. Thus we propose that production of the acid is limited during the winter night while sulphur-bearing precursor gases such as SO_2, following volcanically active periods, or OCS, during background conditions[20], are transported poleward and build up along with O_3. Sulphuric acid may be formed following the oxidation of SO_2 either by OH or O. For relative humidities >1%, OH formed through the photodissociation of O_3

$$O_3 + h\nu \rightarrow O(^1D) + O_2$$

$$O(^1D) + H_2O \rightarrow OH + OH$$

is the dominant oxidizing agent while at lower relative humidities, ground state atomic oxygen, formed through the photodissociation of O_2 and O_3, is dominant[21]. In addition, because at least one of the initial steps in the conversion of OCS to sulphuric acid in the stratosphere involves photolysis

$$OCS + h\nu \rightarrow CO + S$$

it appears likely that sunlight reaching the polar stratosphere

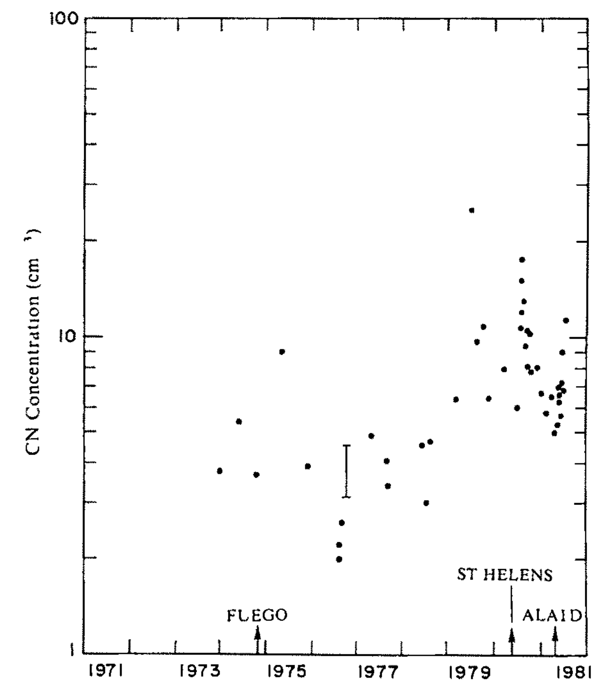


Fig. 4 The condensation nuclei concentration at 20 km as a function of time since measurements began at Laramie in 1973. The vertical bar is the range of concentration observed in four soundings over a 12-day period. The arrows indicate major volcanic eruptions which perturbed the stratospheric $r \geq 0.15$ µm particle concentration.

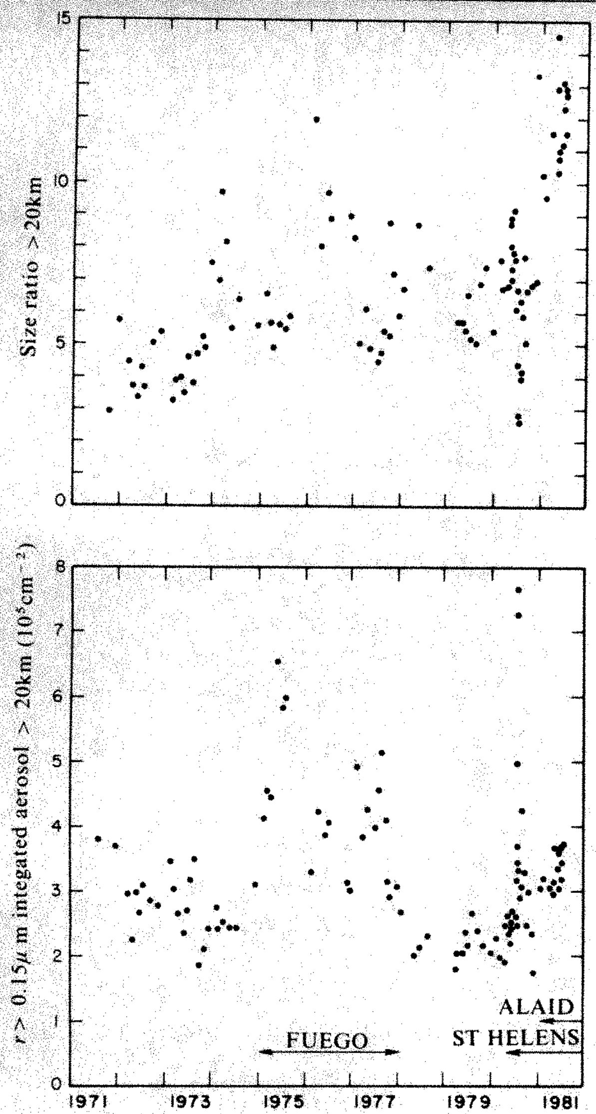


Fig. 5 $r \geq 0.15$ μm particle loading and the ratio of $r \geq 0.15$ μm to 0.25 μm particle loading above 20 km as a function of time since measurements began at Laramie in 1971. Periods when the stratospheric aerosol was perturbed by volcanic eruptions are indicated.

in spring would enhance the production of H_2SO_4 and allow ion complexes to grow rapidly into CN and subsequently be transported to middle latitudes beginning in March.

As the formation of CN in this model is related to multi-ion complexes, it may be related to solar activity which modulates the atmospheric ion density through modulation of the galactic cosmic radiation intensity, large solar flare accelerated particle events, and enhanced auroral activity or precipitation of energetic electrons during geomagnetic activity.

To investigate these speculations, measurements of MIC, CN and larger aerosol at 30 km are required over a solar cycle. Our measurements of CN date back to 1973 and of larger aerosol ($r \geq 0.15$ μm) to 1971. Balloon altitudes in the 27–29 km range were generally attained. Volcanic effects can be large in the 20-km region especially for the larger aerosol. CN are also affected by volcanic eruptions, but these effects are comparatively shortlived due to rather rapid coagulation of these small particles with themselves and with the larger aerosol. While the CN detector functioned satisfactorily only to ~25 km before 1979, CN data at 20 km are reliable and relatively free of volcanic perturbations.

Thus, of the data at hand, it would appear most fruitful to examine the time variation of the CN concentration at 20 km for solar cycle effects. These data are shown in Fig. 4. There seems to be an increasing trend which was evident even before the eruption of Mt St Helens which severely disturbed stratospheric CN concentrations from June to August 1980 at 20 km. The higher concentrations seem to occur at the time of maximum solar activity. In fact, the increasing trend may simply be the result of an accumulation of high altitude events over the past several years. Most of the increase has taken place since early 1979; sunspot maximum occurred in December 1979.

Thus there seem to be more particles at solar maximum, at least for the CN size range. We can investigate the size aspect for the larger aerosol from $r \geq 0.15$ μm measurements. By measuring two sizes ($r \geq 0.15$ and 0.25 μm), we can calculate a size ratio (ratio of the concentrations of the two sizes) which will vary as the size distribution varies. Statistics are poor above 25 km, but if one integrates the concentration above 20 km, a reasonable set of data can be obtained. Figure 5 shows integrated concentration or particle loading (particles cm^{-2}) above 20 km for the $r \geq 0.15$ μm particles and the ratio of loading for $r \geq 0.15$ μm to $r \geq 0.25$ μm for all soundings which reached altitudes in excess of 27 km.

The period 1975–77 was severely disturbed at 20 km by the volcanic eruption of Fuego[22] and the period after May 1980 by the Mt St Helens eruption[17] and the Alaid eruption in April 1981[18]. The period 1978–early 1980 appears to have been free from volcanic eruption.

Sunspot numbers were about 40 in 1973 compared with 140 in 1979 so that the 1972–74 data would represent solar inactive conditions and the 1978–80 data an active solar period. The data in Fig. 5 do not indicate sizable trends between these periods. The size ratio suggests a general trend towards smaller particles (a higher ratio) during the active period. This would accord with the observed increase in condensation nuclei over the period and thus refute Dickinson's hypothesis that ionization control of CN formation could be through the galactic cosmic radiation.

While volcanic effects seem to muddle the interpretation, they may also be used to clarify stratospheric conditions for particle growth between periods. A noticeable effect has been that particles which formed in the stratosphere following the eruption of Fuego in 1974 were substantially larger than those which formed following the eruption of La Soufriere and Sierra Negra in 1979, St Helens in 1980 and Alaid in 1981. The ratio of $r \geq 0.15$ μm to $r \geq 0.25$ μm particle concentrations at the stratospheric maximum was typically 3–4 following the Fuego eruption, 5–6 at background in 1978–79 and 6.5–7.5 following the La Soufriere–Sierra Negra, Mt St Helens and Alaid eruptions. Thus the data are consistent with smaller particles forming during solar maximum periods.

The parameters which control the size to which particles grow in the stratosphere are the amounts of H_2SO_4 and H_2O vapours available and the number of condensation sites present. While the Fuego eruption may have injected relatively more vapours than the other eruptions, it is of interest that the background CN levels were considerably higher at the time of the recent eruptions compared with the time of the Fuego eruption. Higher levels of condensation sites would result in generally smaller particles due to increased competition for the available sulphuric acid and water vapour. Thus, if increased ionization activity during solar maximum results in enhanced levels of ion complexes which form CN, then one would expect volcanically created particles to be smaller during solar maximum, as has been observed.

Conclusions

If solar activity directly affects the stratospheric aerosol, then this influence is probably most readily observable in the number of CN present in the upper stratosphere so that continued monitoring of this important constituent at the highest altitudes is imperative. It appears that a possible mechanism, which

relates ionization produced during solar active periods to the production of CN, now exists and can be used as a basis for further study of this possibly climatic related phenomenon.

We thank the Atmospheric Physics staff and students including J. Anderson, J. Harder, N. Kjome, D. Martell, G. Olson and J. Vali for assistance with the numerous balloon soundings. This work was supported jointly by the NSF (Atmospheric Sciences Section), NASA (Ames Research Center), the NOAA (Aeronomy Laboratory) and the Office of Naval Research (Atmospheric Science Program).

Received 29 December 1981; accepted 16 March 1982.

1. Herman, J. R. & Goldberg, R. A. *Sun, Weather and Climate*, NASA SP-426 (NASA, Washington DC, 1978).
2. McCormac, B. M. & Seliga, T. A. *Solar-Terrestrial Influences on Weather and Climate* (Reidel, Boston, 1979).
3. Ney, E. P. *Nature* **183**, 451 (1959).
4. Roberts, W. O. & Olson, R. H. *Rev. Geophys. Space Phys.* **11**, 731–740 (1973).
5. Dickinson, R. E. *Bull. Am. met. Soc.* **56**, 1240–1248 (1975).
6. Rosen, J. M., Hofmann, D. J. & Singh, S. P. *J. atmos. Sci.* **35**, 1304–1313 (1978).
7. Turco, R. P., Hamill, P., Toon, O. B., Whitten, R. C. & Kiang, C. S. *J. atmos. Sci.* **36**, 699–717 (1979).
8. Rosen, J. M. & Hofmann, D. J. *J. appl. Met.* **16**, 56–62 (1977).
9. Rosen, J. M., Hofmann, D. J. & Kaselau, K. H. *J. appl. Met.* **17**, 1737–1740 (1978).
10. Mohnen, V. A. *Proc. 4th CIAP Conf.*, 478–491 (US Department of Transportation, DOT-TSC-OST-75-38, 1975).
11. Arnold, F., Krankowsky, D. & Marien, K. H. *Nature* **267**, 30 (1977).
12. Ferguson, E. E. *Geophys. Res. Lett.* **5**, 1035–1038 (1978).
13. Arnold, F. & Henschen, G. *Nature* **257**, 521 (1978).
14. Arnold, F. *Nature* **284**, 610–611 (1980).
15. Arnold, F. & Henschen, G. *Geophys. Res. Lett.* **8**, 83–86 (1981).
16. Rosen, J. M. & Hofmann, D. J. *Rep. No. AP-61* (Department of Physics and Astronomy, University of Wyoming, 1981).
17. Hofmann, D. J. & Rosen, J. M. *Rep. No. AP-63* (Department of Physics and Astronomy, University of Wyoming, 1981).
18. Hofmann, D. J. & Rosen, J. M. *Geophys. Res. Lett.* **8**, 1231–1234 (1981).
19. Newell, R. E. *Scient. Am.* **224**, 32–42 (1971).
20. Hofmann, D. J. & Rosen, J. M. *J. atmos. Sci.* **38**, 168–181 (1981).
21. Friend, J. P., Barnes, R. A. & Vasta, R. M. *J. phys. Chem.* **84**, 2423–2436 (1980).
22. Hofmann, D. J. & Rosen, J. M. *J. geophys. Res.* **82**, 1435–1440 (1977).

Expression of two proteins from overlapping and oppositely oriented genes on transposable DNA insertion element IS*5*

Bodo Rak, Monika Lusky* & Majka Hable

Institüt für Biologie III der Universität Freiburg, Schänzlestrasse 1, D-7800 Freiburg, FRG

IS5, the 1,195 base pair transposable DNA element of Escherichia coli, *contains two genes. One gene, which encodes a protein with an apparent molecular weight of 12,300, is, together with its promoter, completely contained within a larger gene which encodes a protein of molecular weight 37,000. The two genes are transcribed in opposite directions.*

TRANSPOSABLE DNA elements have been found in prokaryotic and higher cells (for recent reviews, see refs 1–3). Although there is a large body of information on structural and genetic properties of these elements, the process of transposition itself remains to be elucidated. The differences in the structural organization of various transposable elements may be an indication of differences in at least some of the reaction steps involved in transposition. Several models of transposition propose a two-step reaction[4,5]. In the first step, a transposition intermediate with co-integrated donor and recipient DNAs is formed (provided both molecules involved are circular DNA molecules). Co-integrated molecules have been detected in the cases of Tn*3*[6], IS*1*[7], IS*903*[8] and several others. In the second step, the co-integrated intermediate is resolved. General recombination or, as has been shown for Tn*3*[6], site-specific recombination in recombination-deficient (*recA*) cells, can perform this step. This pathway of transposition is now widely accepted for Tn*3*, although it is still unclear for other elements. In the 5,000 base pair (bp) element Tn*3*, the region carrying the functions required for transposition is at least 3,800 bp long and contains two genes plus an internal resolution site (IRS) for site-specific recombination[6,9–11]. Tn*501* and Tn*1721*[12,13] appear to be similarly organized and, together with Tn*3* and several others, constitute one class of large transposable elements. A different model for transposition, not requiring a co-integrated structure as an intermediate, has recently been proposed for the transposable element *mu* (the 38,000 bp genome of bacteriophage Mu) on the basis of electron microscopic observations[14].

An additional class of transposable elements is represented by the insertion (IS) elements, the first transposable elements to be discovered in *Escherichia coli*[15,16]. They are considerably smaller (800–1,500 bp) and do not carry any detectable phenotype although they are autonomously transposable. The mechanism of transposition of IS elements remains to be elucidated, but their genetic organization is expected to differ from that of the large transposable elements of the Tn*3* type mainly because their coding capacity is restricted. Recent determinations of the DNA sequences of several insertion elements (IS*1*, IS*2*, IS*4*, IS*5*, IS*903* and IS*102*)[8,17–23] have revealed an open reading frame extending over most of the elements' lengths. A corresponding protein has only been detected for IS*4*[24]. Interestingly, all these elements also possess a second, shorter reading frame of significant length and totally contained within the first one. This reading frame is in opposite orientation to the longer one and has the same codon–codon register. Here we present evidence for the simultaneous expression of two proteins from both opposing reading frames on IS*5* under the control of element-coded transcription and translation start signals.

Two promoters on IS*5* precede the large and small reading frames

Analysis of the DNA sequence of insertion element IS*5*[20–22] revealed two overlapping reading frames in opposite orientations. If both reading frames represent functional genes they must be preceded by start signals for transcription and translation. To test IS*5* for potential promoters we cloned different IS*5* DNA fragments in front of a structural gene whose gene product could easily be measured and which did not possess a promoter. Plasmid vector pKO1 (ref. 25), which carries the galactokinase structural gene without a preceding promoter sequence, was used. Readthrough of ribosomes from any gene that may be inserted proximal to the galactokinase gene is prevented in this vector by nonsense codons in all three reading frames. DNA sequences inserted into restriction sites in front of the galactokinase gene could thus be tested for promoter

* Present address: Department of Molecular Biology, University of California, Berkeley, California 94720, USA.

0028–0836/82/190124—05$01.00

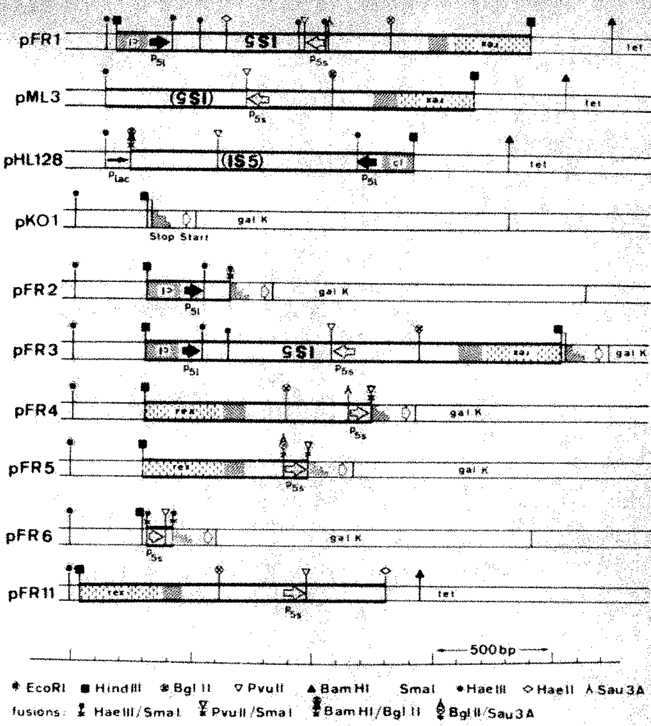


Fig. 1 Relevant structures and restriction sites of plasmids used. Note that the right arm of IS*5* is defined as the one carrying the *Eco*RI recognition site. A 1,720 bp *Hin*dIII fragment including the entire IS*5* element[32] was prepared from λKH100:IS*5* DNA[26] and ligated into the *Hin*dIII site of the vector pML2 to yield pFR1. Plasmid vector pML2 is a derivative of pBR322 in which a DNA fragment from position 1,095 to 2,485 (ref. 46) is deleted (ref. 47 and M.L., unpublished). Plasmid pML3 was obtained by *in vitro* deletion of the small *Eco*RI fragment of pFR1. The construction of pHL128 has been described earlier[32]. Plasmid pKO1 (ref. 25) consists of a segment of pBR322 DNA from the *Pvu*II restriction site at position 2,067 (which is destroyed) to the *Eco*RI site at position 4,361 plus the galactokinase gene. Single restriction sites for *Eco*RI, *Hin*dIII and *Sma*I are located proximal to the galactokinase gene, followed by translation stop codons in all three reading frames ( ). A translation start sequence precedes the galactokinase gene (⇒). There is no proximal promoter sequence that could give significant transcription of the galactokinase gene. Plasmid pFR2 was constructed by substituting the 17 bp *Hin*dIII–*Sma*I fragment of pKO1 with the 357 bp *Hin*dIII–*Hae*III fragment of pFR1 containing the right end of IS*5* up to the first *Hae*III site (IS*5*: 1,195–980, IS*5* sequence coordinates as in refs 20, 21). Plasmid pFR4 was similarly constructed, but the *Hin*dIII–*Pvu*II fragment of pFR1 containing the left end of IS*5* (IS*5*: 1–517) was used as the substituting fragment. For the construction of pFR3, pKO1 was cut with *Hin*dIII, treated with calf intestinal alkaline phosphatase and ligated with the *Hin*dIII fragment of pFR1, which contained the entire IS*5* element. To construct pFR5, plasmid pFR1 DNA was digested with *Hin*dIII, *Bgl*II and *Pvu*II and the resulting fragments were separated on a 1.2% agarose gel. The 585 bp *Hin*dIII–*Bgl*II fragment and the 359 bp *Bgl*II–*Pvu*II fragment were eluted[48], and the 359 bp fragment was digested with *Sau*3A, resulting in two fragments of 251 and 108 bp respectively. The fragments were separated on a 7.5% polyacrylamide gel and the 108 bp *Sau*3A–*Pvu*II fragment was eluted[49]. It was then mixed with the 585 bp *Hin*dIII–*Bgl*II fragment and pKO1 DNA deleted for the 17 bp *Hin*dIII–*Sma*I fragment and the mixture was ligated. Lysates of independent transformants[50] were analysed for the structures of the plasmids present using *Sau*3A digests as well as *Hin*dIII–*Taq*I and *Hin*dIII–*Hae*II double digests. To construct pFR6, a 975 bp *Hin*dIII–*Bst*EII fragment (which contains the left end of IS*5* up to IS*5*:548; the *Bst*EII site lies 8 bp to the right of the *Hae*III site at IS*5*:540, not marked in the figure) was isolated from pFR1 DNA. This DNA fragment was digested with *Hae*III and the resulting fragments separated on a 7.5% acrylamide gel and eluted. The 105 bp fragment bordered by the *Hae*III sites at IS*5*:435 and IS*5*:540 was ligated into pKO1, cut with *Sma*I and treated with calf intestinal alkaline phosphatase. Small lysates from individual transformants were analysed by double digestion with *Eco*RI and *Pvu*II and with *Taq*I and *Pvu*II. Plasmid pFR11 was constructed as follows. (1) The IS*5*-containing *Hin*dIII fragment of pFR1 was inverted by digestion of pFR1 DNA with *Hin*dIII and re-ligation. (2) The plasmid containing the inversion (pFR8) was digested with *Bst*EII and *Bam*HI, and both resulting fragments separated on a 0.8% agarose gel and eluted. (3) The smaller of the two fragments, containing IS*5*:543–1,195, the λ*cI* part plus pBR322 sequences of pML2 from position pBR322:29 to position 375 (ref. 46), was restricted with *Hae*II and fractionated on a 1.8% agarose gel. (4) The *Bst*EII–*Hae*II fragment (IS*5*:543–838) and the *Hae*II–*Bam*HI fragment (pBR322:236–375) were eluted, ligated and then digested with *Bst*EII and *Bam*HI to eliminate head-to-head ligation products. (5) The large *Bst*EII–*Bam*HI fragment containing the pML2 moiety (see step 2) was added together with additional ligase. Small lysates were made from individual transformants and identified by restriction with endonucleases *Bam*HI, *Bst*EII and *Eco*RI. (6) One plasmid (pFR10) carried a deletion of pFR8 from the *Hae*II site at IS*5*:838 to the *Hae*II site at pBR322 coordinate 236 of vector pML2. (7) Since pFR10 carries a deletion of the C-terminal part of the tetracycline gene, the small *Eco*RI–*Bam*HI fragment of pBR322 was substituted for the small *Eco*RI-*Bam*HI fragment of pFR10 giving rise to pFR11. The structure of pFR11 was verified by analysis with restriction endonucleases *Eco*RI, *Bst*EII, *Pvu*II, *Hin*dIII, *Bam*HI, *Pst*I and *Hae*II. Plasmid pFR11 contains the promoter, translation start and 104 amino acids of the IS*5* small reading frame plus leader and coding sequence of the larger tetracycline gene which, when transcribed and translated, would give a protein of 450 amino acids.

activity. Figure 1 shows the pertinent structures of vector pKO1 and its derivatives harbouring different IS*5* DNA fragments. Table 1 shows the Gal phenotypes on indicator plates of *galK*$^-$ strains harbouring these plasmids, as well as the amounts of galactokinase expressed by the various plasmids *in vivo* and in a cell-free system. Nucleotide map positions are those given in ref. 20. Orientation of IS*5* is defined with respect to its integration site in the *cI* gene of bacteriophage λ in mutant λKH100 (ref. 26), and its short *Eco*RI arm (right arm) points towards the beginning of gene *cI*. The source of the different subfragments inserted into pKO1 was pFR1 (Fig. 1), which contains the 1,195 bp IS5 element on a 1,720 bp *Hin*dIII fragment derived from λKH100 DNA.

Plasmid pFR2 contains the right arm of IS*5* up to the *Hae*III site at position IS*5*: 980 inserted in front of the galactokinase gene as a *Hin*dIII–*Hae*III fragment. The data presented in Table 1 show that a promoter of moderate strength active both *in vivo* and in the cell-free system, is located on this cloned fragment. In addition to IS*5* sequences (IS*5*: 980–1,195), the *Hin*dIII–*Hae*III fragment contains an internal segment of the λ*cI* structural gene, λ: 38,150–38,291 (G. Hobom, personal communication), known not to contain any promoter signals[27,28]. We therefore conclude that a promoter, p_{5I}, which transcribes towards the centre of IS*5*, is located on the right arm of the element. Deletion of the hybrid vector insert *Eco*RI fragment from pFR2 results in a Gal$^-$ phenotype on transformation of a *galK*$^-$ strain (pFR7, data not shown). This defines the location of p_{5I} between the *Eco*RI site on IS*5* (IS*5*: 1,097) and the end of the element (IS*5*: 1,195). The localization of a functional promoter in this region agrees with the prediction of a promoter (p_{5I}) at IS*5*: 1,175–1,139 from the sequencing data[20]. This promoter precedes the long open reading frame of IS*5*. Galactokinase activity was significantly reduced when all of IS*5* was inserted proximal to the galactokinase gene in the same orientation as the IS*5* fragment in pFR2 (pFR3). This reduction in expression of *galK* may be due to partial termination of transcription at a terminator sequence, t_{5I}, located on the left arm of IS*5*[20].

Cloning of the left arm of IS*5* (as *Hin*dIII–*Pvu*II fragment) into pKO1 in front of *galK* (pFR4; Fig. 1) resulted in synthesis of galactokinase in a cell-free system but not *in vivo* (Table 1). This rather low but significant promoter activity should be located in the IS*5* segment of this fragment, because the λ segment also present, which contains the end of the λ*cI* gene and the beginning of the λ*rexA* gene (λ: 37,727–38,153; G. Hobom, personal communication) should not code, in this orientation, for any promoter reading towards the galactokinase gene[27,28]. Deletion of the DNA between the *Bgl*II site at IS*5*: 158 and the *Sau*3A site at IS*5*: 409 within pFR4, giving rise to pFR5 (Fig. 1), did not abolish the observed promoter activity (Table 1).

The position of this promoter was mapped more precisely by constructing hybrid plasmids each carrying one of the different fragments that can be generated by *Hae*III digestion of the 975 bp DNA segment between the *Hin*dIII site at λ: 37,727 and the *Bst*EII site at IS*5*: 548. The fragments were inserted in front of the galactokinase gene in both possible orientations. Only one of the *Hae*III fragments (IS*5*: 435–540) was capable of promoting galactokinase synthesis in the cell-free system when inserted into pKO1 (pFR6, Table 1). Its

Table 1 Activity of promoter signals on IS*5*

Plasmid	Gal phenotype	Galactokinase (units) Cell free	*In vivo*
pKO1	−	13.0	5.4
pFR2	+	82.4	91.5
pFR3	(+)	29.5	11.8
pFR4	−	35.2	1.5
pFR5	−	38.5	2.3
pFR6	−	46.6	4.8

E. coli strain N100 (*galK, recA*) was transformed with the different plasmids indicated and Gal phenotypes were detected on McConkey galactose ampicillin agar plates. N100/pFR2 colonies were immediately red on these plates whereas N100/pFR3 developed red colonies only after 1 day incubation at 37 °C. All other plasmids gave white colonies in N100. Galactokinase synthesis directed by the various plasmids was also measured in strain N100. Bacteria were grown to $A_{546} = 1.0$ in minimal salt medium supplemented with casamino acids and glucose, collected by centrifugation and frozen. Cell-free synthesis of galactokinase directed by the different plasmid DNAs was carried out as described by Zubay *et al.*[42] with modifications[43]. Synthesis was for 45 min with 40 μg ml^{-1} of plasmid DNA. Galactokinase was assayed from the frozen cells or from the cell-free reaction mixture as described elsewhere[25,43,44]. Galactokinase activity is expressed in nmol of ^{14}C-galactose phosphorylated per minute and per 1.0 A_{546} bacteria (*in vivo*) or per ml reaction volume (in the cell-free system). Galactokinase synthesis directed by plasmid pKO1 alone may be explained by readthrough from promoter p_4 of pBR322, which is the most efficient promoter present on pBR322, located around position 2,900 in clockwise orientation[45]. Data given represent the average of three or more independent measurements with variations not exceeding 3 units for the *in vivo* assays or 5 units for the cell-free assays.

orientation in this clone is the same relative to *galK* as to the small reading frame in IS*5* (Fig. 1). This DNA fragment overlaps with the beginning of the IS*5* small reading frame. Its promoter activity can, therefore, be correlated with a DNA fragment immediately preceding the IS*5* small reading frame. Other hybrid plasmids directed galactokinase synthesis in this system to the level of pKO1 or below (data not shown).

The 105 bp *Hae*III DNA fragment in pFR6 contains the sequence CAAATTG (IS*5*: 479–485) which qualifies as a Pribnow box and, in the −35 region, the sequence CTTGGTAGCC[29,30]. We suggest that this promoter, which we have called p_{5s}, is the transcription start signal for the IS*5* small gene. The position of the p_{5s} promoter is thus different from its localization on the basis of sequence analysis (ref. 20). There are two possible translation start codons for the IS*5* small reading frame at IS*5*: 495 and IS*5*: 525, coding for proteins of 118 and 108 amino acids respectively. The location of the Pribnow box at IS*5*: 479–485 suggests that translation starts at IS*5*: 525, where a translation start codon is preceded by a Shine–Dalgarno[31] sequence. The absence of activity of p_{5s} *in vivo* may indicate that its activity is negatively regulated, possibly by the other IS*5* elements present in the host strain (see below).

The positions of both promoters have been confirmed by the electron microscopic analysis of transcription complexes with purified *E. coli* RNA polymerase (P. Langridge, personal communication).

Expression of two proteins from opposite strands of IS*5* DNA

Plasmids containing all or part of IS*5* DNA (Fig. 1) were used to direct the incorporation of ^{35}S-methionine into proteins synthesized in a cell-free protein-synthesizing system and in minicells (Fig. 2*A*). In both systems, pFR1, which contains the entire IS*5* element, directed the synthesis of a protein with an apparent molecular weight of 37,000 (Fig. 2*A*, lanes *b, f*). This protein was not detectable when the parental vector DNA (pML2) was used as template (lanes *a, e*). Another plasmid (pML3), containing only the long *Eco*RI arm of IS*5* (IS*5*: 1–1,097), directed the synthesis of the pML2 vector proteins only (Fig. 2*A*, lanes *c, g*). This is consistent with the analysis of IS*5* sequence data, which leads to the prediction that the NH_2-terminus and promoter of the IS*5* large gene are located on the small *Eco*RI arm of IS*5* (IS*5*: 1,097–1,195) with direction of transcription from right to left.

Correlation between the IS*5* large gene of 326 codons and the 37,000 molecular weight (MW) protein is further supported by the data obtained using pHL128 (ref. 32) as template. In pHL128, a large IS*5* fragment from the intact right end to the penultimate codon of the large gene at position IS*5*: 158 on the left side has, by *Bgl*II–*Bam*HI fusion, been linked to a foreign DNA fragment containing the *lac*UV5 promoter sequence. This fusion eliminates the carboxy terminus of the IS*5* protein and the reading frame is extended into the *lac* promoter fragment. The resulting fusion protein should contain 344 amino acids. As predicted, pHL128 directed the synthesis of a new protein with an apparent molecular weight of 39,000, about 2,000 larger than the original IS*5* protein (Fig. 2*A*, lanes *d, h*).

When ^{3}H-leucine, the most abundant amino acid in the IS*5* small gene coding sequence (18/108 amino acids[20–22]), was used for protein labelling using pFR1 and pML3 as templates, a second protein with an apparent molecular weight 12,300 was detected. This protein was not observed with the parental vector pML2 as template (Fig. 2*B*). The λDNA sequences present in pFR1 and pML3, as well as the different vector–insert fusion sequences in these hybrid plasmids, do not contain any reading frame long enough to account for the 12,300-MW protein (G.

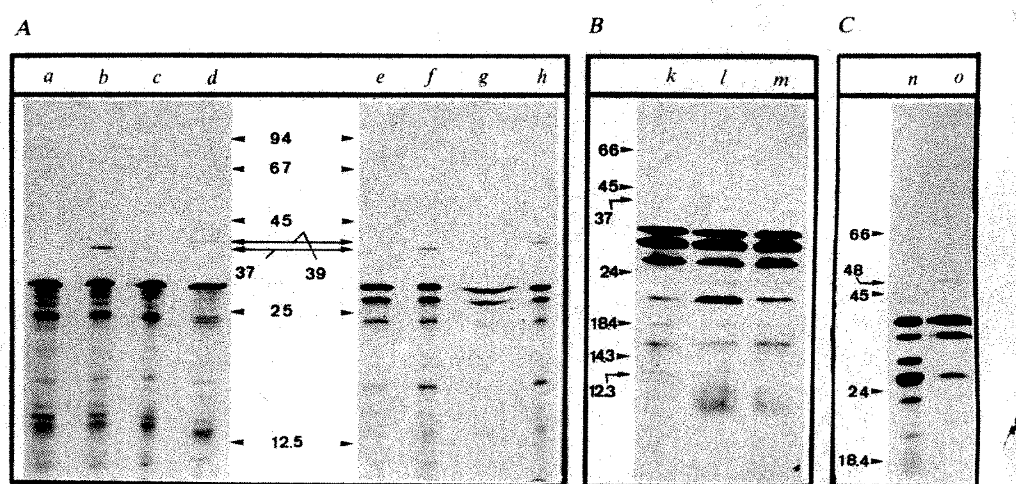


Fig. 2 Expression of proteins encoded by IS*5* in a cell-free protein-synthesizing system and in minicells. *A*, results obtained with ^{35}S-methionine. Cell-free protein synthesis was carried out as described in Table 1 legend with ^{35}S-methionine (1,007 Ci mmol^{-1}; 20 μCi per 100 μl assay). Minicell-producing strain P678-54 was transformed with the different plasmid DNAs, inoculated 1:50 into LB medium containing ampicillin (30 μg ml^{-1}) and grown overnight. Isolation and labelling of minicells were performed as described previously[24] using 0.5 A_{600} units of minicells and 40 μCi ^{35}S-methionine (1,007 Ci mmol^{-1}) in a total volume of 0.5 ml. Proteins (5×10^5 c.p.m. per track) were separated on 12.5% SDS-polyacrylamide gels[51] and the gels were dried and autoradiographed. The position of standard protein size markers are indicated ($\times10^3$). Lanes *a–d*: cell-free synthesis; lanes *e–h*: minicell synthesis. *a, e*, pML2; *b, f*, pFR1; *c, g*, pML3; *d, h*, pHL128. *B*, minicells were prepared as above and labelled with ^{3}H-leucine (65 Ci mmol^{-1}, 130 μCi per assay). Proteins (4×10^4 c.p.m. per track) were separated on 17% SDS-polyacrylamide gels which were fluorographed at −70 °C (ref. 52). Lane *k*, pML3; *l*, pFR1; *m*, pML2. *C*, minicells were prepared and assayed as described above with 110 μCi ^{3}H-leucine (65 Ci mmol^{-1}) and 110 μCi ^{3}H-arginine (53 Ci mmol^{-1}). Proteins (8×10^4 c.p.m. per track) were fractionated on 12.5% SDS-polyacrylamide gels and the radioactive bands visualized by fluorography. Lane *n*, pBR322; *o*, pFR11. The high background (seen previously[24]) is due to the low specific activity of the tritiated amino acids and the necessity to optimize the minicells' protein-synthesizing capacity.

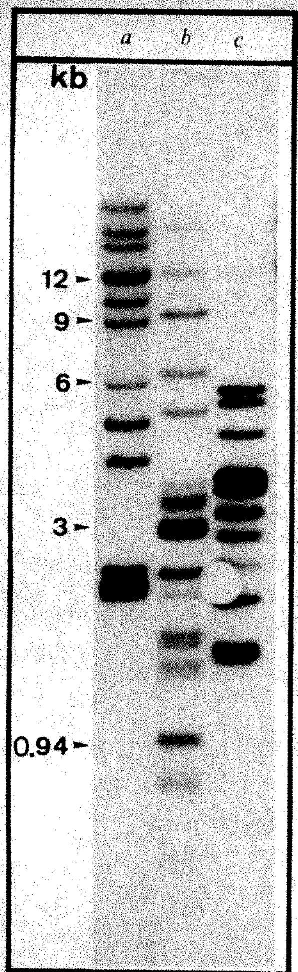


Fig. 3 Southern analysis of IS*5* sequences in *E. coli* K-12 strain 431 Δ (*chl*D–*nad*A). *E. coli* DNA (5 μg) was digested with restriction endonucleases *Bgl*II (*a*), *Pvu*II (*b*), *Hin*dII (*c*) and electrophoresed on a 1% horizontal agarose gel. The fractionated DNA fragments were transferred to nitrocellulose by the method of Southern[37] and hybridized to the radioactively labelled IS*5* internal *Bgl*II–*Eco*RI fragment. Note that *Pvu*II cuts once within IS*5*, between the *Bgl*II and the *Eco*RI site.

Hobom and M. Kröger, personal communication) and we conclude that this protein is the product of the IS*5* small gene. The 12,300-MW protein was produced in the presence and absence of expression of the IS*5* large gene (Fig. 2*B*, lanes *k*, *l*) and is, therefore, not a degradation product of the 37,000-MW protein.

Further support for expression of the IS*5* small gene in minicells was obtained by gene fusion studies. In plasmid pFR11 (Fig. 1), the left arm of IS*5* DNA up to the fourth codon from the end of the small gene was fused, in frame, to the coding sequence of the large tetracycline-resistance gene of pBR322. The resulting fusion protein would be expected to have a molecular weight of 48,800 if the IS*5* small protein translational start is used. In accordance with this prediction, a 48,000-MW protein was encoded by pFR11 in minicells provided with tritiated leucine and arginine (which comprise 118 (59+59) out of 450 amino acids of the fusion protein) (Fig. 2*C*).

IS*5* is present in multiple copies in *E. coli* K-12 genome

Most of the insertion elements found in the *E. coli* genome seem to be present in multiple copies (IS*1*[33], IS*2*[34] and IS*3*[35]), with the exception of IS*4* of which only a single copy has been found in most *E. coli* strains tested[36]. Hybridization data using the gel blotting technique described by Southern[37] revealed 10–12 copies of IS*5* in the DNA of the *E. coli* K-12 strain used (Fig. 3). This strain is a derivative of W8, a prototrophic *E. coli* K-12 from J. Weigle's collection. Interestingly, the same number of IS*5* copies has recently been determined for four different *E. coli* K-12 strains kept separately for many years[21,53].

Discussion

Our data on the functional analysis of IS*5* are summarized in Fig. 4.

From the DNA sequence of IS*5*, two overlapping genes can be proposed: a small one of 324 bp which is completely contained within a large gene of 978 bp. These genes have identical codon registers although transcription would have to proceed in opposite directions. They are preceded by signal structures for the initiation of RNA and protein synthesis. Our data present evidence that the transcription signal structures are indeed functional: RNA synthesis is initiated from the cloned promoter of the large gene both *in vivo* and in a cell-free system. Both genes are transcribed and translated in minicells and the gene products identified correspond in molecular weight to the values expected from the coding regions of their genes (Fig. 2). Activity of the isolated promoter p_{5s} on the other hand was only observed in the cell-free system and not *in vivo*. If we assume that both IS*5* genes carry functions required for transposition, this observation may indicate a complex regulation of the small gene and, hence, of transposition on the level of transcription. Indeed, regulation of transposition has been observed for transposons Tn*5*[38] and Tn*10*[39] which are flanked by IS elements as well as for Tn*3*[9–11,40]. The data presented here are consistent with the hypothesis that the small gene may require its own product *in vivo* for functioning and/or it underlies negative control mechanisms.

The data provide evidence for the unprecedented case of complete genes being transcribed and translated in an antiparallel fashion from the same nucleotide sequence. The highest degree of freedom for the sequence of coding triplets in two genes evolved in an antiparallel fashion from the same DNA fragment could be achieved by codon/codon register. DNA sequence analysis shows that the two IS*5* genes are arranged in this mode.

Comparison of the available DNA sequence data for seven insertion elements revealed similar sequence organization with overlapping, antiparallel, long and short reading frames in codon/codon register in all[8,17–23] but one[41] case. This arrangement of genetic information may thus possibly be a common feature of IS elements.

We do not understand the selective advantage for a genetic structure with such a highly condensed information content, as shown by IS*5*. Selective pressure may have been favourable for a small size of genetic elements which themselves lack selective genes and face the elimination mechanism of the bacterial host cell. Alternatively, selection for a limited size of the element may lie in some step of the yet unknown mechanism by which insertion elements transpose.

We thank K. McKenney for providing plasmid pKO1 before publication, W. Szybalski for bacteriophage λKH100, M. Kröger and G. Hobom for communicating and discussing results before publication, and P. Langridge and C. F. Beck for carefully reading the manuscript. This work was supported by grant Ra276/2 from the Deutsche Forschungsgemeinschaft.

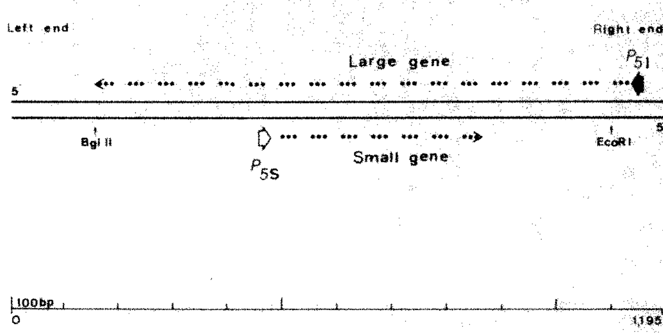


Fig. 4 Schematic presentation of the functional structures on IS*5*. P_{5l} and P_{5s} denote the locations of the promoters functionally detected on IS*5*, preceding the IS*5* large gene and IS*5* small gene respectively. The large arrows indicate the direction of transcription. Dots represent the codon/codon register of the triplets used by the two IS*5* genes.

Received 15 February; accepted 18 March 1982.

1. Starlinger, P. *Plasmid* **3,** 241–259 (1980).
2. Calos, M. P. & Miller, J. H. *Cell* **20,** 579–595 (1980).
3. Kleckner, N. *A. Rev. Genet.* **15,** 341–404 (1981).
4. Grindley, N. D. F. & Sherratt, D. *Cold Spring Harb. Symp. quant. Biol.* **43,** 1257–1262 (1978).
5. Shapiro, J. A. *Proc. natn. Acad. Sci. U.S.A.* **76,** 1933–1937 (1979).
6. Arthur, A. & Sherratt, D. *Molec. gen. Genet.* **175,** 267–274 (1979).
7. Ohtsubo, E., Zenilman, M. & Ohtsubo, H. *Proc. natn. Acad. Sci. U.S.A.* **77,** 750–754 (1980).
8. Grindley, N. D. F. & Joyce, C. M. *Proc. natn. Acad. Sci. U.S.A.* **77,** 7176–7180 (1980).
9. Heffron, F., McCarthy, B. J., Ohtsubo, H. & Ohtsubo, E. *Cell* **18,** 1153–1163 (1979).
10. Gill, R., Heffron, F., Dougan, G. & Falkow, S. *J. Bact.* **136,** 742–756 (1978).
11. Heffron, F., So, M. & McCarthy, B. *Proc. natn. Acad. Sci. U.S.A.* **75,** 6012–6016 (1978).
12. Altenbuchner, J., Choi, C. L., Grinsted, J., Schmitt, R. & Richmond, M. H. *Genet. Res.* **37,** 285–289 (1981).
13. Schmitt, R., Altenbuchner, J. & Grinsted, J. in *Molecular Biology, Pathogenicity and Ecology of Bacterial Plasmids* (ed. Levy, S. B.) 359–370 (Plenum, New York, 1981).
14. Harshey, R. M. & Bukhari, A. I. *Proc. natn. Acad. Sci. U.S.A.* **78,** 1090–1094 (1981).
15. Jordan, E., Saedler, H. & Starlinger, P. *Molec. gen. Genet.* **102,** 353–363 (1968).
16. Shapiro, J. A. *J. molec. Biol.* **40,** 93–105 (1969).
17. Johnsrud, L. *Molec. gen. Genet.* **169,** 213–218 (1979).
18. Ghosal, D., Sommer, H. & Saedler, H. *Nucleic Acids Res.* **6,** 1111–1122 (1979).
19. Klaer, R., Kühn, S., Tillmann, E., Fritz, H. J. & Starlinger, P. *Molec. gen. Genet.* **181,** 169–175 (1981).
20. Kröger, M. & Hobom, G. *Nature* **297,** 159–162 (1982).
21. Schoner, B. & Kahn, M. *Gene* **14,** 165–174 (1981).
22. Engler, J. A. & Van Bree, M. P. *Gene* **14,** 155–163 (1981).
23. Bernardi, A. & Bernardi, F. *Nucleic Acids Res.* **9,** 2905–2911 (1981).
24. Trinks, K., Habermann, P., Beyreuther, K., Starlinger, P. & Ehring, R. *Molec. gen. Genet.* **182,** 183–188 (1981).
25. McKenney, K. *et al.* in *Gene Amplification and Analysis, Vol. II: Structural Analysis of Nucleic Acids* (eds Chirikjian, J. G. & Papas, T. S.) 383–415 (Elsevier, New York, 1981).
26. Blattner, F. R., Fiandt, M., Hass, K. K., Twose, P. A. & Szybalski, W. *Virology* **62,** 458–471 (1974).
27. Szybalski, W. in *Regulatory Biology* (eds Copeland, J. C. & Marzluff, G. A.) 3–45 (Ohio State University Press, 1977).
28. Jones, M. O., Fischer, R., Herskowitz, I. & Echols, H. *Proc. natn. Acad. Sci. U.S.A.* **76,** 150–154 (1979).
29. Rosenberg, M. & Court, D. *A. Rev. Genet.* **13,** 319–354 (1979).
30. Siebenlist, U., Simpson, R. B. & Gilbert, W. *Cell* **20,** 269–281 (1980).
31. Shine, J. & Dalgarno, L. *Proc. natn. Acad. Sci. U.S.A.* **71,** 1342–1346 (1974).
32. Lusky, M., Kröger, M. & Hobom, G. *Cold Spring Harb. Symp. quant. Biol.* **45,** 173–176 (1980).
33. Nyman, K., Nakamura, K., Ohtsubo, H. & Ohtsubo, E. *Nature* **289,** 609–612 (1981).
34. Saedler, H. & Heiss, B. *Molec. gen. Genet.* **122,** 267–277 (1973).
35. Deonier, R., Hadley, R. & Hu, S. *J. Bact.* **137,** 1421–1424 (1979).
36. Klaer, R. *et al. Cold Spring Harb. Symp. quant. Biol.* **45,** 215–224 (1980).
37. Southern, E. *J. molec. Biol.* **98,** 503–517 (1975).
38. Biek, D. & Roth, J. R. *Cold Spring Harb. Symp. quant. Biol.* **45,** 189–191 (1980).
39. Beck, C. F., Moyed, H. & Ingraham, J. L. *Molec. gen. Genet.* **179,** 453–455 (1980).
40. Chou, J., Casadaban, M. J., Lemaux, P. & Cohen, S. N. *Proc. natn. Acad. Sci. U.S.A.* **76,** 4020–4024 (1979).
41. Auerswald, E. A., Ludwig, G. & Schaller, H. *Cold Spring Harb. Symp. quant. Biol.* **45,** 107–113 (1980).
42. Zubay, G., Chambers, D. A. & Cheong, L. C. in *The Lactose Operon* (eds Beckwith, J. R. & Zipser, D.) 375–391 (Cold Spring Harbor Laboratory, New York, 1970).
43. Wetekam, W., Staack, K. & Ehring, R. *Molec. gen. Genet.* **116,** 258–276 (1972).
44. Wilson, D. B. & Hogness, D. S. *Meth. Enzym.* **8,** 229–240 (1966).
45. Stüber, D. & Bujard, H. *Proc. natn. Acad. Sci. U.S.A.* **78,** 167–171 (1981).
46. Sutcliffe, J. G. *Cold Spring Harb. Symp. quant. Biol.* **43,** 77–90 (1978).
47. Lusky, M. & Botchan, M. *Nature* **293,** 79–81 (1981).
48. Wienand, U. & Feix, G. *FEBS Lett.* **98,** 319–323 (1979).
49. Maxam, A. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74,** 560–564 (1977).
50. Birnboim, H. C. & Doly, J. *Nucleic Acids Res.* **7,** 1513–1523 (1979).
51. Laemmli, U. K. *Nature* **227,** 680–685 (1970).
52. Bonner, W. M. & Laskey, R. A. *Eur. J. Biochem.* **46,** 83–88 (1974).
53. Nakamura, K. & Inouye, M. *Molec. gen. Genet.* **183,** 107–114 (1981).

Controlled transcription of a human α-interferon gene introduced into mouse L cells

Ned Mantei & Charles Weissmann

Institut für Molekularbiologie I, Universität Zürich, Hönggerberg, 8093 Zürich, Switzerland

Mouse L cells were transformed with a cloned chromosomal human interferon-α_1 gene, linked to a herpes thymidine kinase gene as selectable marker. In 8 out of 12 cell lines containing the human interferon gene, infection with Newcastle disease virus stimulated synthesis of messenger RNA for the human interferon—in parallel with that of the endogenous mouse interferon genes.

HUMAN interferons are assigned to three major classes[1]: α, β and γ. The α-interferons (IFN-α) are encoded by a gene family consisting of a dozen or more non-allelic members[2–9]. IFN-α preparations derived from induced human leukocytes or lymphoblastoid cell lines consist of five or more[10,11] species, indicating that many of the genes are expressed *in vivo*.

Normal cells usually do not produce interferon and do not contain detectable levels of interferon mRNA[12]. Treatment with inducers such as viruses or double-stranded RNA leads to the transient formation of IFN-α and/or IFN-β mRNA as well as interferon, 4–24 h after induction[1].

We show here that mouse L cells transformed with a cloned chromosomal IFN-α_1 gene produce correctly initiated IFN-α_1 mRNA in parallel with the appearance of mouse interferon mRNA after induction with Newcastle disease virus (NDV), but not in normal growth conditions. Thus the foreign gene seems subject to the normal control mechanisms of the host cell. A preliminary report of this work has been presented at the International Meeting on the Biology of the Interferon System, Rotterdam (21–24 April 1981).

Human IFN-α transcripts with correct 5′ termini are produced in virus-induced transformed mouse cells

The hybrid phage ML-Ch (*Eco*)/Hchr(AH)IFN-35 carries the gene for human IFN-α_1 (ref. 3). A 7.5-kilobase (kb) *Bam*HI fragment from this phage DNA, containing the IFN-α_1 gene, 5.4 kb of 5′-flanking and 1.2 kb of 3′-flanking sequences, was subcloned into the *Bam*HI site of pBR322 (pChr35-BB; S. Nagata, unpublished). The plasmid TK-M2 carries a thymidine kinase (TK) gene from herpes simplex virus type I (ref. 13). The two plasmids were cleaved at their single *Sal*I sites and joined by T_4 DNA ligase to form concatenates. LMTK$^-$ cells were transfected with the concatenates using the calcium phosphate method[14,15]. After 21 days in hypoxanthine/aminopterin/thymidine medium[16], TK$^+$ colonies were picked and expanded.

The IFN-α_1 DNA content of 13 TK$^+$ cell lines (M2/IFN-1 to M2/IFN-13) was estimated by Southern blotting of *Eco*RI-cleaved cell DNA[17,18] and hybridization with ^{32}P-labelled nick-translated IFN-α_1 cDNA[19]. All cell lines of the M2/IFN series except M2/IFN-3 contained IFN-α_1 DNA, ~0.25 to more than 20 copies per haploid genome (data not shown).

The human IFN-α_1 mRNA content of the cellular RNA was measured by an S_1 mapping procedure[20]. A 480-base pair (bp) *Mbo*II–*Bgl*II fragment of the human chromosomal DNA, which spans the 5′ terminus of the mRNA (see Fig. 6 of ref. 3), was 5′-terminally labelled at the *Bgl*II end. The denatured fragment was hybridized to the RNA, digested with S_1 nuclease and analysed by polyacrylamide gel electrophoresis[21]. Correctly initiated RNA is predicted to protect a ^{32}P-labelled fragment of 330 nucleotides[3]. As little as 0.14 ng poly(A)$^+$ RNA from interferon-producing human leukocytes[19] gave detectable

0028–0836/82/190128—05$01.00

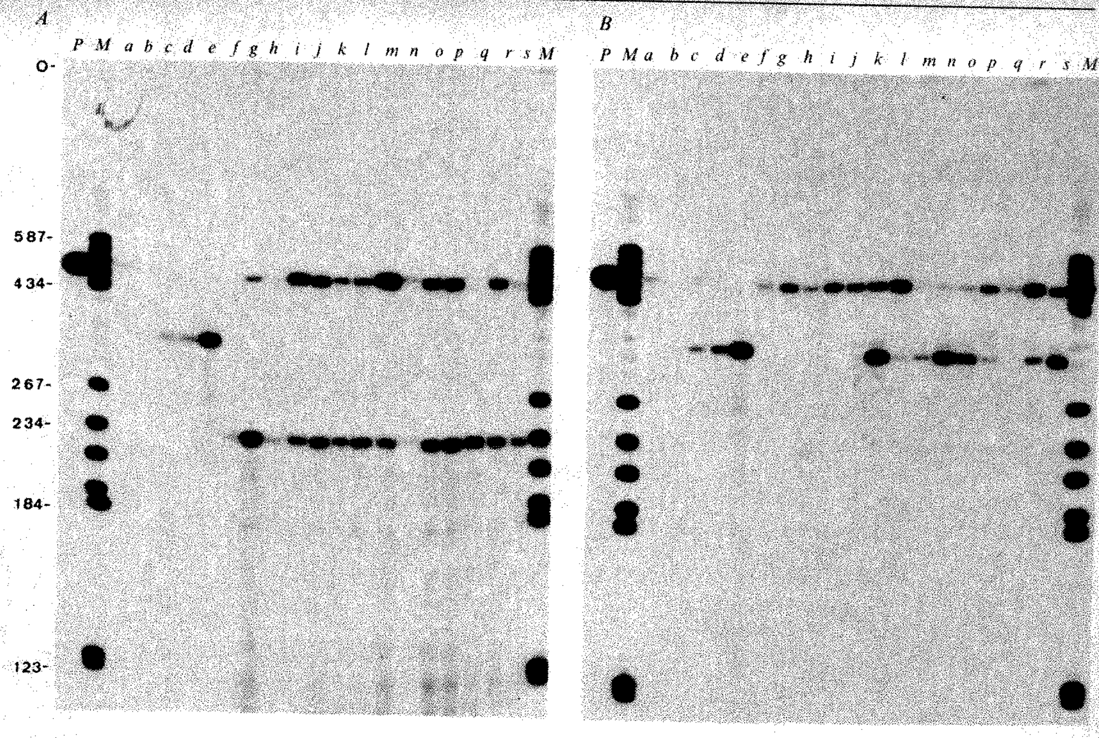


Fig. 1 Detection of human IFN-α_1 RNA with correct 5′ termini in transformed mouse L cells. TK-negative mouse L cells (LMTK$^-$) were transfected with concatenates of the plasmids TK-M2 (ref. 13) and pChr35-BB, which contains a 7.5-kb *Bam*HI fragment from the IFN-α_1 chromosomal DNA[9]. 1.6 µg TK-M2 and 9 µg pChr35-BB were cleaved with *Sal*I and joined with ligase. Ten 9-cm plates of LMTK$^-$ were transfected and TK-positive colonies selected as described previously[15,23]. Control cells were transfected with *Sal*I-cleaved TK-M2 without ligation. RNA was prepared as described in ref. 27. To prepare the probe for S_1 mapping, plasmid pChr35-HBα (ref. 3) was cleaved with *Bgl*II (New England Biolabs) and the 5′ termini labelled with ^{32}P-phosphate[12]. After *Mbo*II cleavage, the 480-bp fragment was isolated. Probe (0.01 pmol, 27,000 c.p.m.) was denatured, hybridized in 80% formamide to 50 µg cell RNA, digested with S_1 nuclease (P-L Biochemicals) and analysed by polyacrylamide gel electrophoresis on a 5% gel, all as described earlier[12,21]. *A*, RNA from uninfected cells. Lanes *a–e*, 0, 0.14, 0.42, 1.25 and 3.75 ng poly(A)$^+$ RNA from interferon-producing human leukocytes[19]. Lane *f*, RNA from cell line M2/S-33, transformed only with TK-M2. Lanes *g–s*, RNA from cell lines M2/IFN-1 to M2/IFN-13, transformed with TK-M2 linked to pChr35-BB. *B*, analysis of RNA from transformed cells infected with NDV. Transformed cells were grown for three generations with hypoxanthine and thymidine but without aminopterin, then infected with an optimal dose of NDV. RNA was extracted after 11 h. Samples as in *A*. Lane *P*, undigested probe. Lane *M*, size marker: pBR322 digested with *Bsp*I and 5′-^{32}P-labelled. Exposure for 21 days at −70 °C with an Ilford intensifying screen.

amounts of this fragment (Fig. 1*A*, lane *b*), whereas 50 µg RNA from transformed cell lines M2/IFN-1 to M2/IFN-13 (Fig. 1*A*, lanes *g–s*) did not. There are therefore fewer than 0.5 human IFN-α_1 mRNA molecules with correct 5′ termini per mouse cell transformed with IFN-α_1 DNA (see below for calculations). The 480-nucleotide fragment corresponds to completely protected probe, and is due, as shown below, to protection by complementary RNA longer than the probe and/or to self-renaturation. RNA from all mouse lines tested, including the line M2/S-33, transformed only with TK-M2 (Fig. 1*A*, lane *f*), gave a labelled fragment apparently 230 nucleotides long. We later found that this fragment and one of ~200 nucleotides (see Fig. 2) appear on electrophoresis of large amounts of undigested probe or undigested probe mixed with nuclease S_1-digested cell RNA (data not shown).

To test whether gene expression could be induced, the transformed cell lines were infected with NDV and RNA prepared 11 h later. As shown by S_1 mapping, RNA from 8 of 13 lines gave the 330-nucleotide fragment (Fig. 1*B*) indicative of correctly initiated human IFN-α_1 mRNA. The level of this mRNA was estimated from the amount of cell RNA per cell (~25 pg)[22], the molecular weight of IFN-α_1 mRNA (~330,000) and the intensities of the 330-nucleotide bands relative to those given by RNA from induced human leukocytes, ~0.4% of which is IFN-α_1 RNA (N.M., unpublished). We estimate that eight of the cell lines contained ~0.5–10 IFN-α_1 transcripts with correct 5′ termini per cell. No transcripts were found in cell line M2/S-33, transformed with TK-M2 alone (Fig. 1*B*, lane *f*).

To examine the kinetics of IFN-α_1 induction, cultures of M2/IFN-5 and of M2/S-34 (transformed with TK-M2 alone) were infected with NDV or mock-infected. After 20 h, media from both infected cultures contained about 4,000 units ml^{-1} of mouse interferon, indicating that mouse interferon was induced in both cell lines. RNA was prepared after 4, 11, 20 and 27 h, and poly(A)$^+$ and poly(A)$^-$ RNA separated by chromatography on oligo(dT)-cellulose[23,24]. IFN-α_1 RNA was assayed by the S_1 mapping procedure described above. In poly(A)$^+$ RNA from infected cultures of M2/IFN-5, correctly initiated IFN-α_1 RNA was present after 11 h (Fig. 2, lane *l*) and, in lower amounts, after 20 h (lane *c*), but not after 4 h (lane *a*) or 27 h (lane *e*). Neither cell line M2/S-34 (lanes *n*, *o*) nor mock-infected M2/IFN-5 (lanes *b*, *d*, *f*, *m*) ever gave a 330-nucleotide fragment. The mobilities of the 330-nucleotide fragments given by M2/IFN-5 RNA (Fig. 2, lane *l*) and by human leukocyte RNA (lane *k*) are the same, implying that the 5′ ends of the RNAs are identical. The signal from poly(A)$^+$ RNA from 150 µg M2/IFN-5 RNA at 11 h after infection (lane *l*) was about 15 times stronger than that from poly(A)$^-$ RNA from 50 µg of the same cell RNA (lane *p*). About 80% of the IFN-α_1 RNA is therefore polyadenylated.

The kinetics of induction of mouse interferon mRNA was followed by injecting samples of poly(A)$^+$ cell RNA into *Xenopus* oocytes and determining antiviral activity on mouse cells. Activity was undetectable at 4 h, reached a maximum at 11 h and dropped to about one-tenth of maximum at 20 h after infection.

The identification of human IFN-α_1 mRNA in virus-induced mouse cells depends on the formation of perfect or almost perfect hybrids between the mRNA and the human DNA probe in stringent hybridization conditions. Human and mouse IFN-α coding sequences differ in about 26% of their residues (G. Shaw and N.M., unpublished), so that only very imperfect hybrids could be formed even in non-stringent hybridization conditions. In fact, no fragments characteristic for correctly initiated human IFN-α_1 mRNA were found when 5′ probes were hybridized to RNA from interferon-producing mouse cells transformed with TK plasmid alone (Fig. 1*B*, lane *f*). The results of the S_1 mapping of RNA from virus-induced mouse cells containing the human IFN-α_1 gene therefore cannot be explained by cross-reaction with mouse interferon mRNA.

Human IFN-α transcripts with incorrect 5′ termini in uninduced, transformed mouse cells are relatively stable

The 3′ ends of the IFN-α_1 transcripts were mapped with a probe spanning the known 3′ terminus. Even after hybridization to only yeast RNA, the S_1-treated probe gave several discrete bands, probably due to incomplete digestion (Fig. 3, lane *a*). Poly(A)$^+$ RNA from induced human leukocytes gave three additional protected fragments of 550, 600 and 780 nucleotides (Fig. 3, lanes *b, c, f*). The 3′ end of the plus-strand IFN-α_1 cDNA in plasmid Z-pBR322 (*Pst*)/HcIF-2h is 554 nucleotides downstream from the *Bgl*II site[12], accounting for the 550-nucleotide fragment. The other protected fragments probably represent other 3′ ends further downstream. Two or more 3′ ends have been found for IFN-α_2 RNA[7] and other RNAs[25,26]. An artefact of the S_1 assay is less likely. RNA from both NDV-infected and mock-infected M2/IFN-5 cells gave the same bands found with induced human leukocyte RNA, albeit at different relative intensities (Fig. 3, lanes *d, e*). This shows that both induced and mock-induced, transformed cells contain human IFN-α_1 transcripts with correct 3′ termini. As no correct 5′ termini had been found in mock-induced cells, we conclude that either the 5′ moiety of these transcripts was degraded or they initiated far upstream of the 'cap' site. In the latter case, the labelled probe used previously to map 5′ ends would be protected over its full length, giving a signal indistinguishable from renatured probe.

To determine whether IFN-α_1 RNA in mock-infected cells originated upstream of the cap site, we used a probe with a 5′-^{32}P-labelled *Eco*RI end at nucleotide 633 in the IFN-α_1 gene, extending upstream to nucleotide −675 in the 5′-flanking region and continuing 103 nucleotides into the adjoining pBR322 DNA (see Fig. 4). All '3′ halves' of IFN-α_1 RNA found in the experiment of Fig. 3 will be detected by this probe, as the *Eco*RI site lies 372 nucleotides downstream from the labelled *Bgl*II end used for the 3′-end mapping. Transcripts initiated further upstream than position −675 will yield a protected fragment of ~1,375 nucleotides which is easily distinguished from undegraded probe (~1,480 nucleotides). Figure 4*B*, lane *e*, shows that RNA from NDV-infected M2/IFN-5 yielded as the major product the 700-nucleotide fragment expected for correctly initiated RNA, with lesser amounts of 1,100- and 1,375-nucleotide fragments. RNA from mock-infected M2/IFN-5 protected fragments of 1,100 and 1,375 nucleotides only (Fig. 4*B*, lane *f*). Mapping of 3′ ends of RNA from the same batch showed that the 5′ ends detected account for most or all of the 3′ ends (data not shown).

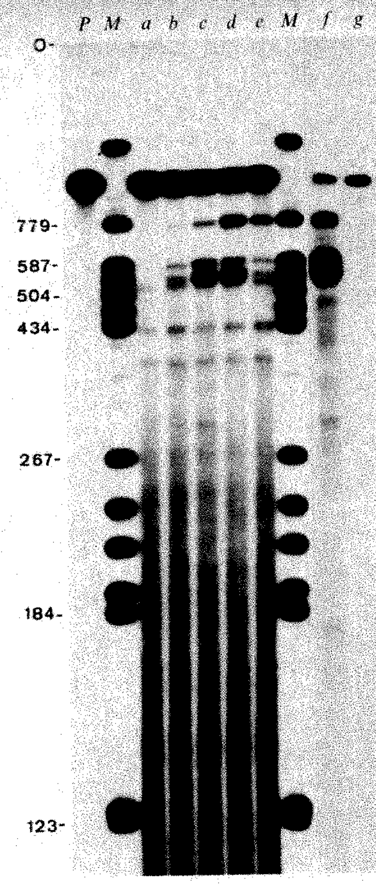


Fig. 3 S_1 mapping of 3′ ends of human IFN-α_1 RNA. The plasmid pChr35-HBα containing the human chromosomal IFN-α_1 sequence was incubated with *Bgl*II and then with T_4 DNA polymerase (P-L Biochemicals) at 15 °C in the presence of 0.1 mM dCTP[35]. After phenol extraction and ethanol precipitation, the termini were 'filled in' with the Klenow fragment of DNA polymerase (Boehringer) in the presence of 0.2 mM each dCTP, dGTP and TTP, and 2.4 µM [α-^{32}P]dATP (NEN). Following cleavage with *Bsp*I, the resulting 1,000-bp *Bgl*II–*Bsp*I fragment (see map in ref. 3) was isolated. S_1 analysis as for Fig. 1. Lanes *a–c, f, g*, 0, 1.2, 3.6, 300 and 0 ng poly(A)$^+$ RNA from interferon-producing human leukocytes[19]. Lanes *d, e*, poly(A)$^+$ RNA (from the preparation described in Fig. 2 legend) derived from 120 µg cell RNA of cell line M2/IFN-5, either 11 h after NDV-infection (*d*) or mock infection (*e*). Lanes *P* and *M* as in Fig. 1. Exposure for 25 days, except that lanes *f* and *g* were exposed for 16 h.

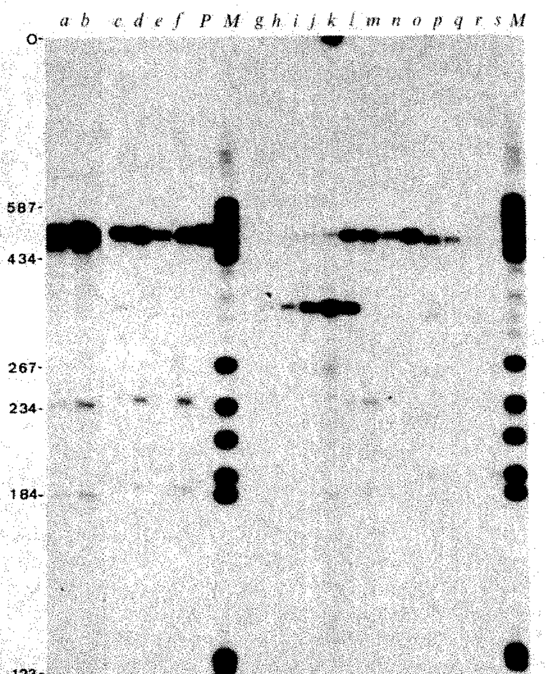


Fig. 2 5′ termini of human IFN-α_1 RNA in transformed mouse cells at different times after infection with NDV. M2/IFN-5 and M2/S-34 cells (see Fig. 1 legend) were infected with NDV or were mock-infected. RNA was prepared at 4, 11, 20 and 27 h after the end of the adsorption period and fractionated by chromatography on oligo(dT)-cellulose[23,24] (P-L Biochemicals, type 7). Poly(A)$^+$ RNA derived from 150 µg cell RNA and poly(A)$^-$ RNA from 50 µg cell RNA were analysed as for Fig. 1. Lanes *g–k*, 0, 0.16, 0.5, 1.5 and 4.5 ng poly(A)$^+$ RNA from interferon-producing human leukocytes[19]. Lanes *a, c, e* and *l*, poly(A)$^+$ RNA from M2/IFN-5 at 4, 20, 27 and 11 h after NDV infection. Lanes *b, d, f* and *m*, poly(A)$^+$ RNA from M2/IFN-5 at 4, 20, 27 and 11 h after mock infection. Lanes *n, o*, poly(A)$^+$ RNA prepared 11 h after infection (*n*) or mock infection (*o*) of M2/S-34. Lanes *p–s*, poly(A)$^-$ RNA prepared 11 h after infection (*p, r*) or mock infection (*q, s*) of M2/IFN-5 (*p, q*) and M2/S-34 (*r, s*). Lanes *P* and *M* as in Fig. 1. The autoradiograph was exposed for 25 days.

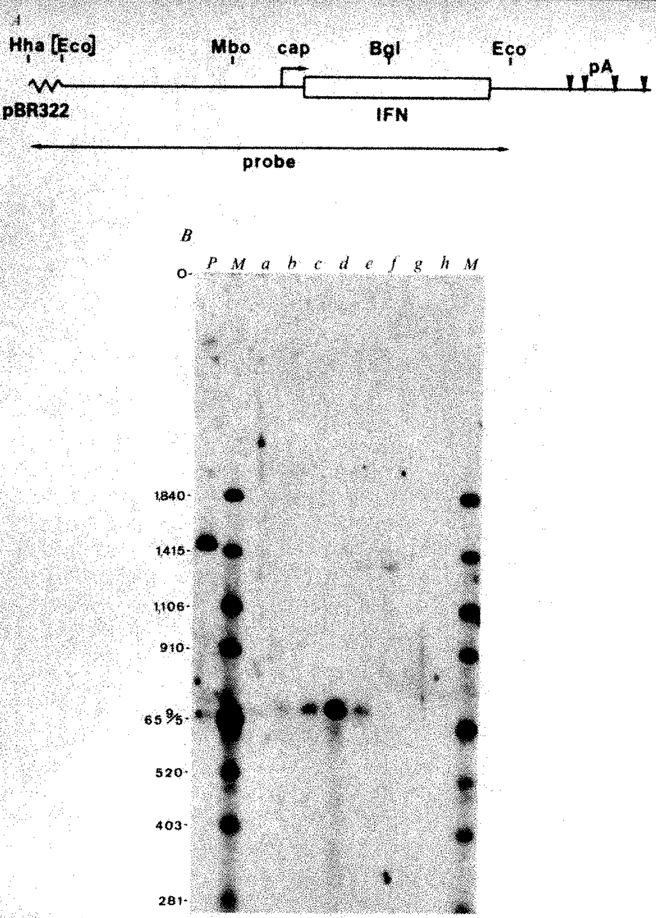


Fig. 4 S_1 mapping of 5′ ends of human IFN-α_1 RNA in transformed mouse cells. M2/IFN-5 and M2/S-34 cells were infected with NDV or mock-infected, and poly(A)$^+$ RNA prepared after 11 h. *A*, the plasmid pChr35-675 is a derivative of pChr35-HBα (ref. 3) in which the *Eco*RI fragment between −675 and the *Eco*RI site of pBR322 was deleted and the new *Eco*RI site destroyed [*Eco*]. pChr35-675 was cleaved with *Eco*RI and the 5′ termini labelled. After cleavage with *Hha*I and *Pst*I, the 1,480-bp *Hha*I-*Eco*RI fragment was isolated. *B*, mapping of 5′ ends. The fragment described above (0.01 pmol, 40,000 c.p.m.) was hybridized to poly(A)$^+$ RNA derived from 125 μg cell RNA, digested with S_1 nuclease at 24 °C and the products analysed by electrophoresis through a 1.6% alkaline agarose gel[36]. Lanes *a*–*d*, 0, 0.42, 1.25 and 3.75 ng poly(A)$^+$ RNA from interferon-producing human leukocytes. Lanes *e*, *f*, RNA from cell line M2/IFN-5 after infection (*e*) or mock infection (*f*). Lanes *g*, *h*, RNA from cell line M2/S-34 after infection (*g*) or mock infection (*h*). Lane *P*, undigested probe; lanes *M*, pBR322 cleaved with *Hin*cII and *Pvu*II, plus pBR322 cleaved with *Alu*I, both terminally labelled. Exposure for 14 days.

The S_1 mapping of poly(A)$^+$ RNA from induced, transformed mouse cells, from the *Bgl*II site to the 5′ and 3′ termini and from the *Eco*RI site to the 5′ termini, gives signals identical to those found with poly(A)$^+$ RNA from induced human leukocytes. We conclude that the induced, transformed mouse cells contain complete human IFN-α_1 mRNA which has the correct termini and is polyadenylated.

Conclusions

Mouse L cells, transformed with a cloned human IFN-α_1 gene, produce a low level of incorrectly initiated interferon transcripts. Most, if not all, of these transcripts initiate upstream of the cap site. Transcripts initiated upstream of the cap site have also been found in the case of the rabbit β-globin gene and the chicken ovalbumin gene introduced into mouse L cells[27,28]. When mouse L cells transformed with the IFN-α_1 gene are infected with NDV, correctly initiated, polyadenylated IFN-α_1 transcripts appear, at levels of 0.5–10 strands per cell. We believe this to be due to *de novo* synthesis, and not to processing of the pre-existing overlong transcripts: if such processing were accidental, it is unlikely that newly formed 5′ termini would coincide precisely with the 5′ termini of natural IFN-α_1 mRNA. On the other hand, it is unlikely that, in physiological conditions, overlong transcripts give rise to mature interferon RNA by processing, because human leukocytes contain no IFN-α-related transcripts before induction[12]. The time course of appearance and disappearance of the human IFN-α_1 RNA in mouse cells is similar to that of the mouse interferon mRNA in the same cells, suggesting a common control mechanism. Whereas mouse interferon was easily measured in the medium, human IFN-α_1 could not be detected, probably because the interferon mRNA level was low, and the specific antiviral activity of IFN-α_1 on human cells is lower by an order of magnitude than that of other human interferons[29].

Induction of interferon mRNA in normal cells could be explained in two ways: (1) although interferon mRNA is synthesized at a similar rate in non-induced and induced cells, transcripts are rapidly degraded in non-induced cells but stabilized in induced cells. (2) There is no (correct) transcription of interferon genes in non-induced cells, and induction activates transcription. The fact that non-induced transformed mouse cells contain about as many incorrectly initiated IFN-α_1 mRNA transcripts as induced cells suggests that, unless a correct 5′ terminus is required for degradation, the stability of interferon transcripts is similar in both cases. Thus, induction would not suppress interferon mRNA degradation, but rather stimulate correct transcription of IFN-α genes, which does not occur to a measurable extent in non-induced cells.

The interferon gene flanked by 5.4 kb of 5′- and 1.2 kb of 3′-chromosomal DNA seems to contain sufficient information to respond to physiological induction, at least to some extent. We estimate that there is about 50 times less human than mouse interferon mRNA in an induced, transformed mouse cell line containing two IFN-α_1 genes per haploid genome; this may be due in part to the fact that the cells contain many more mouse interferon genes, probably more than 10 copies per haploid genome. In addition, there may be some incompatibility between the human and mouse system, and/or effects due to incorrect positioning of the transplanted genes in the mouse genome.

Recently, other cloned genes introduced into host cells have been found to respond to induction—the cloned mouse mammary tumour virus genome[30,31], the rat α2 globulin gene[32] and a *Drosophila* heat-shock gene[33]. Thus, several systems now provide an approach to the identification of the DNA sequences responsible for control of transcription.

This work was supported by the Schweizerische Nationalfonds (3.481.79) and the Kanton of Zürich. We thank S. Nagata and U. Weidle for plasmids, C. Arnheiter for NDV, M. Streuli for injecting oocytes, P. Dierks and J. Fujisawa for [γ-^{32}P]ATP, and C. Arnheiter and J. Fujisawa for helpful advice.

Note added in proof: Ohno and Taniguchi[34] have found that mouse FM3A cells transformed with the human IFN-β gene produce IFN-β and increased amounts of IFN-β mRNA after treatment with NDV or poly(I) · poly(C). Similar results have been obtained by H. Hauser *et al.* (personal communication).

Received 31 December 1981; accepted 16 March 1982.

1. Stewart, W. E. II *The Interferon System* (Springer, New York, 1979).
2. Streuli, M., Nagata, S. & Weissmann, C. *Science* **209**, 1343–1347 (1980).
3. Nagata, S., Mantei, N. & Weissmann, C. *Nature* **287**, 401–408 (1980).
4. Goeddel, D. V. *et al. Nature* **290**, 20–25 (1981).
5. Nagata, S., Brack, C., Henco, K., Schamböck, A. & Weissmann, C. *J. Interferon Res.* **1**, 333–336 (1981).
6. Lawn, R. M. *et al. Science* **212**, 1159–1162 (1981).
7. Lawn, R. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 5435–5439 (1981).
8. Owerbach, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 3123–3127 (1981).
9. Brack, C., Nagata, S., Mantei, N. & Weissmann, C. *Gene* **15**, 379–394 (1981).
10. Allen, G. & Fantes, K. H. *Nature* **287**, 408–411 (1980).
11. Rubinstein, M. *et al. Archs Biochem. Biophys.* **210**, 307–318 (1981).
12. Mantei, N. *et al. Gene* **10**, 1–10 (1980).

13. Wilkie, N. M. et al. Nucleic Acids Res. 7, 859–877 (1979).
14. Graham, & van der Eb, A. J. Virology 52, 456–467 (1973).
15. Wigler, M., Pellicer, A., Silverstein, S. & Axel, R. Cell 14, 725–731 (1978).
16. Szybalski, W., Szybalska, E. H. & Ragni, G. Monogr. natn. Cancer Inst. 7, 75 (1962).
17. Southern, E. M. J. molec. Biol. 98, 503–517 (1975).
18. Wahl, G. M., Stern, M. & Stark, G. R. Proc. natn. Acad. Sci. U.S.A. 76, 3683–3687 (1979).
19. Nagata, S. et al. Nature 284, 316–320 (1980).
20. Berk, A. J. & Sharp, P. A. Cell 12, 721–732 (1977).
21. Weaver, R. F. & Weissmann, C. Nucleic Acids Res. 7, 1175–1193 (1979).
22. Parsons, J. T., Coffin, J. M., Haroz, R. K., Bromley, P. A. & Weissmann, C. J. Virol. 11, 761–774 (1973).
23. Mantei, N., Boll, W. & Weissmann, C. Nature 281, 40–46 (1979).
24. Aviv, H. & Leder, P. Proc. natn. Acad. Sci. U.S.A. 69, 1408–1412 (1972).
25. Tosi, M., Young, R. A., Hagenbüchle, O. & Schibler, U. Nucleic Acids Res. 9, 2313–2323 (1981).
26. Nevins, J. R. & Darnell, J. E. Cell 15, 1477–1493 (1978).
27. Dierks, P., van Ooyen, A., Mantei, N. & Weissmann, C. Proc. natn. Acad. Sci. U.S.A. 78, 1411–1415 (1981).
28. Breathnach, R., Mantei, N. & Chambon, P. Proc. natn. Acad. Sci. U.S.A. 77, 740–744 (1980).
29. Streuli, M. et al. Proc. natn. Acad. Sci. U.S.A. 78, 2848–2852 (1981).
30. Hynes, N. E., Kennedy, N., Rahmsdorf, U. & Groner, B. Proc. natn. Acad. Sci. U.S.A. 78, 2038–2042 (1981).
31. Buetti, E. & Diggelmann, H. Cell 23, 335–345 (1981).
32. Kurtz, D. T. Nature 291, 629–631 (1981).
33. Corces, V., Pellicer, A., Axel, R. & Meselson, M. Proc. natn. Acad. Sci. U.S.A. 78, 7038–7042 (1981).
34. Ohno, T. & Taniguchi, T. Nucleic Acids Res. 10, 967 (1982).
35. Challberg, M. D. & Englund, P. T. Meth. Enzym. 65, 39–45 (1980).
36. McDonnel, M. W., Simon, M. N. & Studier, F. W. J. molec. Biol. 110, 119–146 (1977).

LETTERS TO NATURE

Dynamical features in the northern hemisphere of Saturn from Voyager 1 images

Garry E. Hunt, David Godfrey, Jan-Peter Müller & Robert F. T. Barrey

Laboratory for Planetary Atmospheres, Department of Physics and Astronomy, University College London, London WC1E 6BT, UK

Results of the analysis of Voyager measurements of divergence and inferred vertical velocity of convective cloud features in the northern hemisphere of Saturn are presented here. The winds in this region have been measured by tracking the cloud elements observed in sequences of images. The derived zonal winds imply that these convective features reside at the maximum of strong easterly jets centred at 41° N planetocentric latitude and that the jet is barotropically unstable. The divergences and local dynamics of the Saturn features are found to differ from those of the jovian equatorial plumes.

During the encounter of Saturn by Voyager 1, Smith *et al.*[1] reported observations of distinct cloud features occurring in the northern hemisphere of the planet's atmosphere. In the region bounded by 35–42° N, several bright V-shaped features have been seen, while at ~29° N, a distinct cloud oval is apparent (Fig. 1, feature 3.) (All latitudes in this article are planetocentric.) The large spot is particularly evident in the UV images and was seen to persist throughout the Voyager 1 observational period from 24 August to 10 November 1980. The V-shaped features varied greatly in shape during the observation period but persisted in their longitudinal distribution as they drifted in an easterly jet centred at 41° N. The changes in the visible appearance of these features suggests that they are convective clouds where the rapid variations in the shape and structure of these cloud systems is associated with local vertical motions. There is no evidence from the IR observation of Hanel *et al.*[2] that these features affect the temperature structure of the upper troposphere. However, the IRIS field of view is large compared with the visible cloud structures observed by the imaging system[3]. The IRIS instrument essentially measures the average temperature for the total region seen in the imaging frame. Indeed, any upper atmosphere aerosol layers present would further reduce any possible temperature contrasts.

In the atmosphere of Jupiter, we have shown that the equatorial plumes are convective cloud structures[3,4] and their meteorological characteristics have been estimated by procedures developed using the University College London (UCL) Interactive Planetary Image Processing System[3]. We apply these techniques for the analysis of the Saturn observations and present here the first measurements of divergence of saturnian cloud systems. Using this information, we then calculate the vertical velocities which relate to the observed changes in the cloud structures. In addition, we measure the wind speeds of the clouds in the latitude band 30–45° N and estimate the stability of the flow for the period of the observations. These new measurements provide valuable insight into the variations of the cloud structures and through comparison with our jovian studies[4] this will provide information on the differences in the observed meteorologies.

For this study, a series of blue filter narrow angle images of Saturn was used. The observations were made for the period 3–8 November 1980 when the spatial resolution varied between 120 and 60 km pixel^{-1}.

The images were first radiometrically decalibrated at the Jet Propulsion Laboratory (JPL) using the standard procedures[5]. To provide an accurate navigation of the images that transforms the line and sample system into a planetocentric latitude/longitude coordinate system, two independent procedures were used. At JPL, the method used simultaneous wide and narrow angle frames to locate the centre of the planet from which the subsequent transformation is then determined[5]. An interactive procedure, developed at UCL, uses a 'least-squares' minimization method to fit an ellipse to the image, constrained by the spacecraft navigation and camera data. We have shown elsewhere[6] that both of these methods produce similar results although the error in determining the position of the centre of the planet becomes greater than 1 pixel when 20% or less of the planet's limb is available (see ref. 7).

Measurements of velocities of the upper clouds in the Saturn atmosphere have been made by interactively tracking individual features observed in the images. The techniques have been shown[6] to be accurate to within $\pm 3\ \mathrm{m\ s^{-1}}$ in zonal and meridional winds at a spatial resolution of 60 km pixel^{-1} and produce results that compare favourably with other schemes[1]. The velocities are measured relative to the SYSTEM III radio period of 10 h 39.4 min (ref. 8) using images separated by time intervals of 10–20 h. Considerable contrast enhancement of the image was first necessary to show clearly the cloud elements to be tracked.

The measurements of the divergence of the V-shaped clouds were made by the procedures set out in ref. 4. The sequence of 10 blue narrow angle images were first map projected into a Mercator format centred at 41° N using a single pixel interpolation scheme developed at UCL[6]. Then, the divergence of these features was estimated from the changing cloud areas. We estimate the area of the convectively active region from the assumption that the brightness in the cloud system, which we have previously shown[4,9,10] to be a reliable approach.

If we now assume that the vertical motion in the layer of outflow takes place over a vertical scale D, then the vertical velocity required to produce the observed divergence field is:

$$W \leq D\ \mathrm{div}\ \mathbf{V} \qquad (1)$$

We estimate D as the scale height of the atmosphere at the cloud top, which will then enable an estimate to be made of the vertical motions associated with these cloud systems.

0028-0836/82/190132—03$01.00

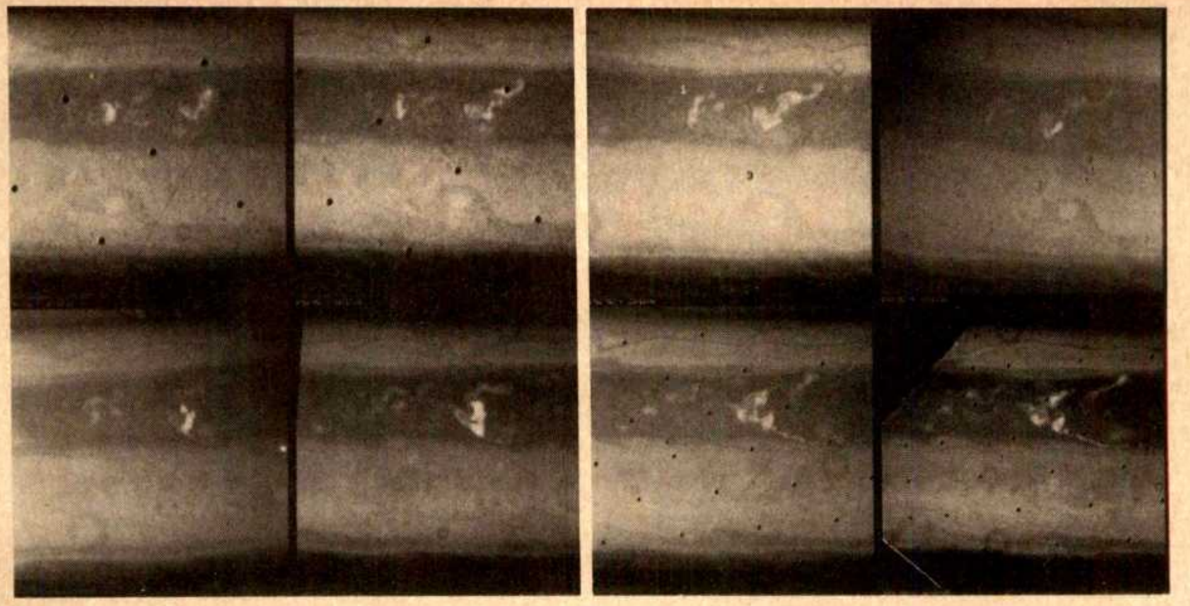


Fig. 1 Sequence of eight blue Mercator narrow angle images of the mid-latitudes of Saturn. The dates and times of the images are: 15:05:22, 3 November 1980; 11:57:22, 4 November 1980; 22:11:46, 4 November 1980; 08:25:22, 5 November 1980, 20:46:58, 5 November 1980; 17:22:10, 6 November 1980; 12:08:34, 8 November 1980; 22:24:34, 8 November 1980.

Figure 1 shows the variation in the V-shaped cloud structures for the period 15:05:22, 3 November to 22:24:34, 8 November 1980. All times are GMT at the spacecraft. This time sequence clearly indicates the large morphological change in the convective cloud structures. At the edge of feature 2, considerable developments in the clouds become apparent in the final three images of the sequence. Figure 1 shows less noticeable changes during this time.

The third main cloud feature seen in Fig. 1 is the UV spot. The prominence of this feature at short wavelengths is consistent with it being higher than its surrounding clouds. A well defined flow pattern can be seen to the west of this feature in Fig. 1, which is morphologically similar to the 'turbulent' regions to the west of the Great Red Spot and White Ovals on Jupiter.

In Fig. 2, we summarize the basic characteristics of the motions in the mid-latitude region bounded by 30–50° N. The mean zonal velocity is shown in Fig. 2*a*; a strong easterly jet is situated at 41° N which corresponds to the apex of the V features. We find also that the strongest shear occurs along the southern edge of this jet. The small inflection in the zonal velocity profile at 42–43° N is associated with small changes in the cloud morphologies.

To discuss the significance of these results and the stability of the flow, we estimate $\bar{u}$, $d^2\bar{u}/dy^2$, and their associated errors, using the procedures set out in ref. 11. The measured zonal flow is composed of contributions from the mean flow, the eddies and errors which arise from effects of spacecraft navigation, preprocessing of the data and uncertainties in the identification of individual features. Following the procedures of Ingersoll *et al.*[11], we estimate the error in $\bar{u}$, $\sigma(\bar{u})$ as ± 5 m s^{-1}. We compute $d^2\bar{u}/dy^2$ at latitudinal position y_n by:

$$(d^2\bar{u}/dy^2)_n = (\bar{u}_{n+2} - 2\bar{u}_n + \bar{u}_{n-2})/\Delta y_n^2$$

where $\Delta y_n = y_{n+1} - y_{n-1}$.

Then the error in $d^2\bar{u}/dy^2$ is $\sim \pm\sqrt{6}\sigma(\bar{u})/\Delta y^2$ (ref. 11).

The stability of the retrograde jet may be investigated by examining the barotropic stability condition for a zonal velocity $\bar{u}$ (see, for example, ref. 12), given by:

$$\beta - d^2\bar{u}/dy^2 \geqslant 0$$

where $\beta = (2\Omega/r) . \cos\theta$, Ω is the planetary rotation, r the planetary radius (local radius or curvature), θ the planetodetic latitude and y the northerly direction. In Fig. 2*b*, we have superimposed the value of β with $d^2\bar{u}/dy^2$, while the additional

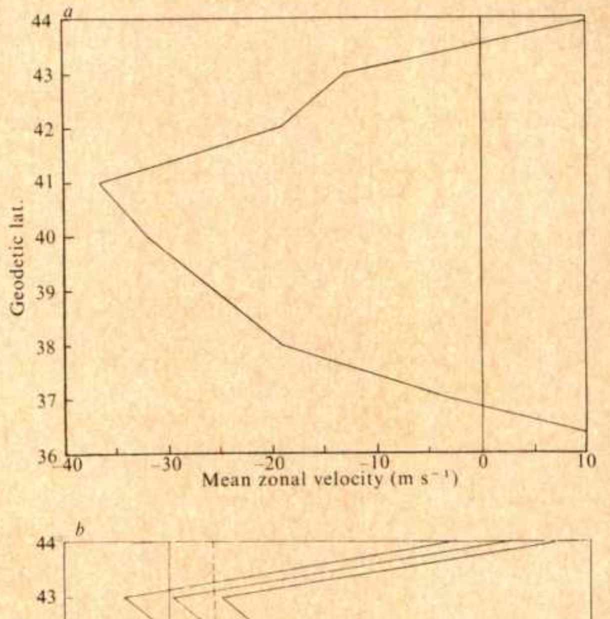


Fig. 2 The dynamical properties of the region 30–50° N derived in this study. *a*, The zonal wind $\bar{u}$ is shown with 1σ r.m.s. values. *b*, The curve for $d^2\bar{u}/dy^2$, together with the value of β at 41° N. The central curve is the variation of $d^2\bar{u}/dy^2$. The curves at the left and right show the variation in $d^2\bar{u}/dy^2$ when the errors are taken into account.

curves illustrate the errors in $d^2\bar{u}/dy^2$ estimated by the procedures set out above. We see that the flow between 40 and 41.5° N is barotropically unstable in more stringent conditions. The figure further suggests that the region extending to 37° N is close to being unstable in these conditions. The V-cloud features reside in the centre of the unstable region, which is also the maximum in the easterly jet. The stable region of the flow lies outside the latitudes of 37–42° N.

Figure 3 shows the divergence of these active regions. There is little change in the characteristics of the UV spot, which is apparent from the image sequence seen in Fig. 1. The changes in area shown in Fig. 2 for the convective features were used to determine the 90% of maximum brightness threshold contour using the procedures described in ref. 4. These variations in area of the active clouds may then be used to estimate the divergence of the features which are shown in Fig. 3. As expected, feature 2 is the most active, with values reaching $\sim 5\times10^{-5}\,s^{-1}$. The other convective feature is less active, with a maximum value of $3\times10^{-5}\,s^{-1}$. A negative divergence implies the dissipation of the cloud feature.

Comparing the results with those of our study[4] of the jovian equatorial plumes, we find that the divergences are much stronger for the saturnian features. The largest Saturn values are more than three times greater than those associated with the most active jovian plume. Assuming that the anvil expansion we are measuring takes place at ~500 mbar level[2] over a depth of a density scale height of 40 km (ref. 13), then the divergence maxima imply vertical motions of $\sim 2\,m\,s^{-1}$.

On Jupiter, these large scale motions are propagated throughout the troposphere[3]. Although the saturnian vertical motions are apparently stronger, the IR observations do not suggest any large scale perturbations in the temperature structure in a horizontal or vertical direction. It is possible that the larger aerosol layers in the troposphere of Saturn provide sufficient opacity to diminish the temperature contrasts. Furthermore, the IRIS observations are essentially the average response for the narrow angle field of view, so that these small scale convective clouds may not produce a sufficiently large signature to be noticeable.

The present estimates of larger vertical motions in the atmosphere of Saturn compared with Jupiter are consistent with the differences in the zonal velocity and internal structure of the two planets. The internal heat source on Saturn supplies approximately twice as much energy to the atmosphere as solar heating[2]. This situation, together with the smaller gravity, which then leads to a larger atmospheric scale height on Saturn, is consistent with the larger convective motions discussed above.

From these first estimates of divergence and vertical motions of cloud features observed in the northern hemisphere of Saturn we find that the associated flows of the convective clouds are stronger than the similar cloud features observed in the atmosphere of Jupiter. The vertical velocities we estimate for these motions are of a similar magnitude to large scale terrestrial convective storms (see, for example, ref. 9), although on Saturn these storms occupy a considerably greater spatial scale.

We find that the Saturn convective cloud features are situated in the maximum of the easterly jet which corresponds to a barotropically unstable region of the atmosphere (Fig. 2). It is possible that these mesoscale cloud features are generated by large scale convergence in the deep cloud layers beneath the visible surface of the planet. These stronger motions are consistent with the differences in the structures of Saturn and Jupiter.

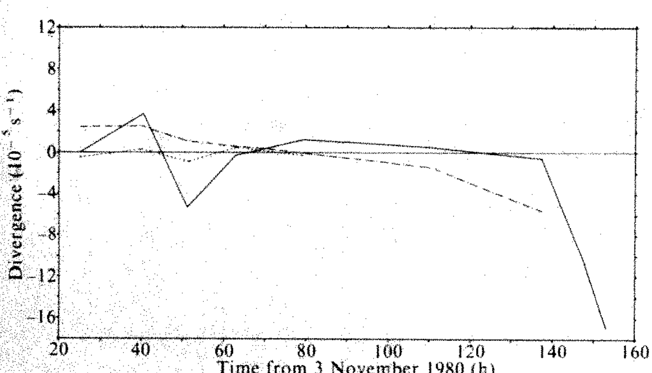


Fig. 3 Measurements of divergence of cloud features in the northern hemisphere of Saturn for 150 h, commencing at 00 h, 3 November 1980. The convective clouds are features 1 (—·—); feature 2 (——); · · · · , the UV spot.

We thank Glenn Garneau, Eric Korsmo and Gary Yagi at JPL for assistance. This work is supported by the Science and Engineering Research Council.

Received 8 March; accepted 1 April 1982.

1. Smith, B. A. *et al. Science* **212**, 183–191 (1981).
2. Hanel, R. A. *et al. Science* **212**, 192–201 (1981).
3. Hunt, G. E., Conrath, B. J. & Pirraglia, J. A. *J. geophys. Res.* **86**, 8777–8783 (1981).
4. Hunt, G. E., Müller, J-P. & Gee, P. *Nature* **295**, 491–494 (1982).
5. Jepsen, P. *et al. J. Br. interplanet. Soc.* **33**, 315–320 (1980).
6. Hunt, G. E. *et al. Phil. Trans. R. Soc.* (in the press).
7. Sromovsky, L. A., Revercomb, H. E., Suomi, V. E., Limaye, S. S. & Krauss, R. J. *J. atmos. Sci.* (submitted).
8. Kaiser, M. J., Desch, M. D., Warwick, J. W. & Pearce, J. B. *Science* **209**, 1238–1240 (1980).
9. Ward, N. R., Hunt, G. E. & Saunders, R. W. *Int. J. Remote Sensing* (in the press).
10. Fraedrich, K., Ruprecht, E. & Tunte, U. *J. appl. Met.* **15**, 1312–1316 (1976).
11. Ingersoll, A. P. *et al. J. geophys. Res.* **86**, 8733–8744 (1981).
12. Holton, J. *An Introduction to Dynamical Meteorology* (Academic, New York, 1973).
13. Newburn, R. & Gulkis, S. *Space Sci. Rev.* **3**, 179–271 (1973).

Enhancement of the number of muon catalysed fusions

L. Bracci & G. Fiorentini

Istituto di Fisica dell'Università di Pisa, and INFN, Sez. di Pisa, 56100 Pisa, Italy

We discuss here the effect of an external electric field on the motion of the $(\mu\alpha)$ ions which are produced in the reaction $(dt\mu)\rightarrow(\mu\alpha)+n$. We find that appreciable enhancements of the $(\mu\alpha)$ stripping probability can be achieved for strong fields. This effect can be useful for enhancing the number of nuclear fusions which a muon can catalyse in a deuterium–tritium mixture.

The possibility of using the muon catalysis of nuclear reactions for the purpose of energy production has been a long-standing dream of particle physicists[1]. Recently, Petrov[2] presented a scheme which, although based on some optimistic assumptions, can yield a positive energy balance. This scheme involves the muon catalysis of $d+t\rightarrow\alpha+n$ reactions and is supported by the recent discovery[3] of a very high formation rate of the d–μ–t molecular ion, $\lambda_{d\mu t}\geq 10^8\,s^{-1}$ at liquid hydrogen density. Indeed, the theoretical analysis of the process, based on the so-called resonant mesomolecule formation[4,5], allows for the possibility of extremely high values of $\lambda_{d\mu t}$.

In the d–t mixture, the most serious limitation to the efficiency of the fusion chain is the possibility that the muon could stick to the α particle produced in the nuclear reaction, the sticking probability being $W=1.2\%$ (refs 6, 7). The kinetic energy of the $(\mu\alpha)$ ion is $T_{in}=3.5$ MeV, which corresponds to a velocity $v_{in}=5.8\,e^2/\hbar$. During the slowing down process there is a chance for the $(\mu\alpha)$ to be stripped, so that the muon can take part again to the fusion chain. At a density $\rho_0=10^{19}$ atom cm^{-3} a fraction $R=0.20$ of the sticked muons are reactivated, so that the effective sticking probability is $W_{ef}=W(1-R)=0.96\%$ while at $\rho=10\rho_0$ one has $R=0.21$ and $W_{ef}=0.95\%$ (see refs 6, 7). The reason for such a limited gain (as depicted in Fig. 1) is that the stripping rate λ_{st} becomes very small as the $(\mu\alpha)$ reaches low velocity, $v\leq e^2/\hbar$. To improve the value

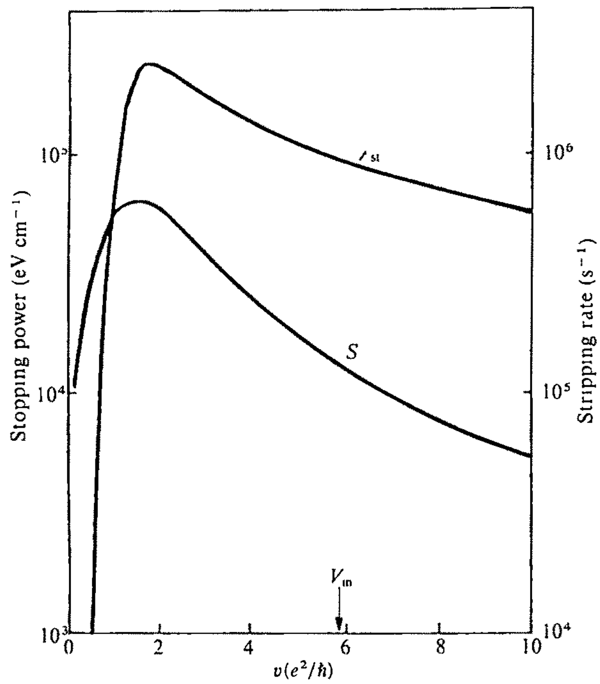


Fig. 1 Stopping power S in eV cm^{-1} and stripping rate λ_{st} in s^{-1} versus $v\hbar/e^2$ at density $\rho = 10^{19}$ atoms cm^{-3}

of the reactivation coefficient R the $(\mu\alpha)$ should be kept as long as possible in the velocity range where the stripping rate is high In principle, this can be achieved by making up for the energy loss due to the stopping power $S(v)$ of the medium by means of an applied electric field

Clearly the strength E of the electric field which is needed for this purpose is of the order $E \simeq S(v_{in})/e$, which means $E \simeq 15$ kV cm^{-1} at density $\rho = \rho_0$ This is a rather large value and consequently one has to keep the target density as low as possible On the other hand, lower limits to the value of the density are required by the condition that the processes relevant to the fusion chain (muon slowing down, atomic capture and atomic cascade, mesomolecule formation) are fast compared with the muon decay rate, $\lambda_\mu = 4\ 5 \times 10^5$ s^{-1} We assume the mesomolecule formation to be fast at $\rho \geq \rho_0$, which is optimistic but consistent with the present experimental observation and needs comparison with future measurements of $\lambda_{dt\mu}$ In this way one gets $C = \rho/\rho_0 \geq 1$ We consider two cases, $C = 1$ and $C = 10$

To calculate the reactivation coefficient we solve numerically the equations of motion of the $(\mu\alpha)$ under the combined action of the electric field and of the dissipative force $S(v)$ This allows us to solve the evolution equation for the number of stripped $(\mu\alpha)$s We first consider the case of a static, uniform electric field E a field that is constant as long as the muon is present in the target and in the region travelled by the $(\mu\alpha)$ One can envisage several configurations depending on the geometry of the apparatus Of course, the value of R is configuration dependent We focus on two extreme cases first where the initial velocity of all the $(\mu\alpha)$s is aligned with the electric field and second where the initial velocity is randomly oriented The results, presented in Fig 2, allow a straightforward interpretation The peak for the parallel case corresponds to the critical value of the field $E/C = 13\ 1$ kV cm^{-1}, when the dissipative force $S(v_{in})$ is exactly balanced by the electric field In this case the $(\mu\alpha)$ velocity is constant and the reactivation probability is $R = \lambda_{st}(v_{in})/(\lambda_{st}(v_{in}) + \lambda_\mu)$ In a higher electric field the $(\mu\alpha)$s are accelerated and therefore the stripping probability is diminished, as is clear from Fig 1 In lower electric fields the $(\mu\alpha)$s are slowed down but it takes longer to reach thermal energy than in the zero-field case and consequently $R(E) > R(0)$ Lower values of R are generally found in the isotropic case because for a fraction of the $(\mu\alpha)$s the slowing down time is shortened by the electric field In any case there are significant enhancements of R for sufficiently high electric fields

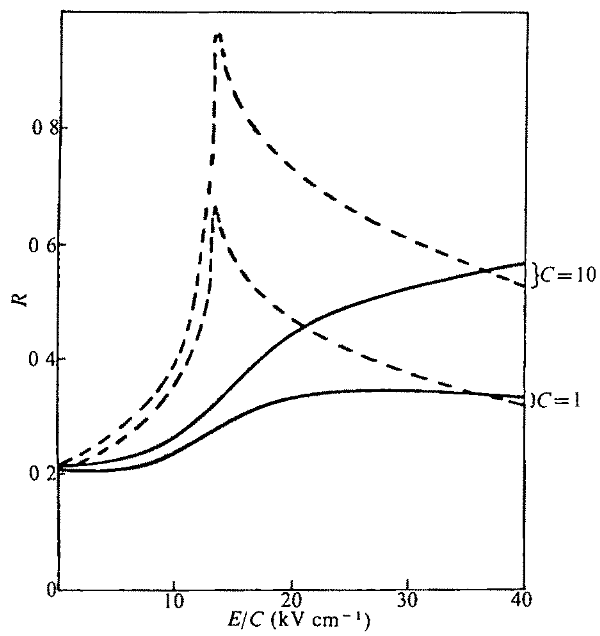


Fig 2 Reactivation coefficient R versus E/C for the parallel case (dashed line) and the isotropic case (solid line)

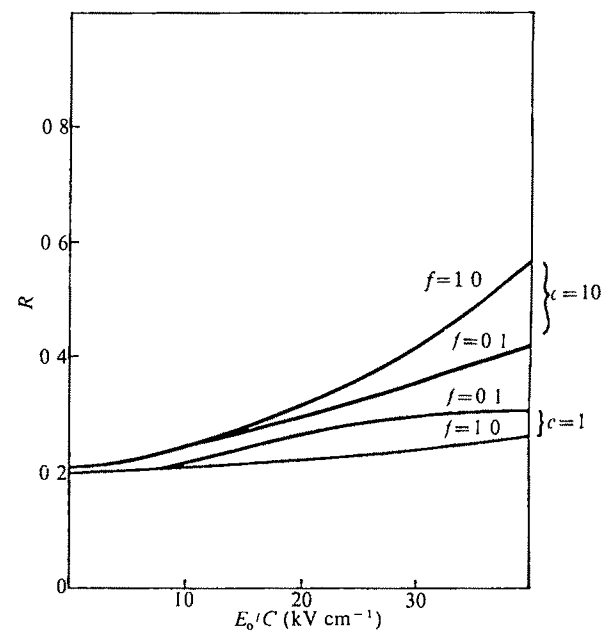


Fig. 3 Reactivation coefficient R versus E_0/C for different values of $f = \omega/\omega_0$ with $\omega_0 = 10^7$ s^{-1}

Another situation of interest is that of an oscillating electric field, $E = E_0 \cos(\omega t + \varphi)$ As the phase (φ) of the field at the time of the $(\mu\alpha)$ formation $(t = 0)$ is random, one needs to take a statistical average over φ for the calculation of the reactivation probability This is equivalent to averaging over the initial orientations of the $(\mu\alpha)$ velocity and consequently one expects the geometry of the apparatus to be irrelevant for the reactivation effect For this reason we only consider the isotropic case The qualitative features of the $(\mu\alpha)$ motion are as follows The velocity is exponentially unstable near the initial value, with a time constant $\tau \simeq M/S'(v_{in})$ where M is the $(\mu\alpha)$ mass and S' is the derivative of the stopping power with respect to the velocity In the range of field strength and frequency of interest (see Fig 3) this results in a slowing down of the $(\mu\alpha)$ The time the $(\mu\alpha)$ spends with $v \geq e^2/\hbar$ is a decreasing function of ω because, as soon as the electric field becomes opposite to v, the

($\mu\alpha$) is quickly decelerated Next, for $\omega > 2\pi\lambda_\mu$ the ($\mu\alpha$) velocity can perform several oscillations around zero, with an amplitude $A = eE_0/(\omega^2M^2 + S'(0)^2)^{1/2}$ before muon decay In this stage, for a strong electric field and a high density, the ($\mu\alpha$) spends some time in a region where the stripping rate is appreciable These considerations allow an easy explanation of the results depicted in Fig 3 The opposite behaviour with respect to ω at $C = 1$ and $C = 10$ arises from the different contributions of the slowing down stage and of the oscillatory regime to the reactivation probability R

We conclude that one can enhance the reactivation efficiency by supplying energy to the ($\mu\alpha$)'s by means of an electric field, provided that it is strong eneough On the other hand severe limitations to the strength of the electric field arise because the spark breakdown, is initiated for $E/C \simeq 3$ kV cm^{-1} (ref 8) Thus pulsed and oscillatory field seem preferable, because there it is possible to reach higher peak values than with static fields

An interesting case would be to use the oscillating electric field in conjunction with a magnetic field B At the cyclotron frequency $\omega_c = eB/Mc$ one gets the same situation as with a static electric field Furthermore, the magnetic field can be useful in inhibiting the spark breakdown

We feel that this approach, despite technical difficulties, can be a help to bring muon catalysed fusion closer to practical applications

We thank F Kottman, L Olivo and, particularly, G Zweig for helpful discussions

Received 14 October 1981, accepted 16 March 1982

1 Alvarez, L W *Advent exp Phys* α, 72 (1972)
2 Petrov Yu V *Nature* **285**, 466 (1980), *Proc XIV LNPI Winter school* Leningrad (1979)
3 Bystritsky, V M *et al*, *Phys Lett* **94B**, 476 (1980)
4 Vesman, E A *Soviet Phys JEPT Lett* **5**, 91 (1967)
5 Gershtein, S S & Ponomarev L I *Phys Lett* **72B**, 80 (1977)
6 Gershtein, S S *et al Zh Eksp Teor Fiz* **80**, 1690 (1981)
7 Bracci L & Fiorentini, G *Nucl Phys* **A364**, 383 (1981)
8 Alston *High-Voltage Technology* Atomic Energy Research Establishment, Harwell

Search for superheavy elements in monazites using chemical enrichment

R. Stakemann, R. Heimann, G. Herrmann, G. Tittel & N. Trautmann

Institut fur Kernchemie, Universitat Mainz, D-6500 Mainz, FRG

Evidence for the existence of superheavy elements in monazite inclusions embedded in Madagascan mica and surrounded by giant radioactive haloes was given by Gentry *et al.*[1] who observed photons with energies corresponding to predicted $L_{\alpha1}$ X-ray energies of element 126 (at 27.25 keV) and also of elements 116, 124 and 127 in irradiations of such crystals with collimated proton beams. For an unambiguous identification, the detection of further members of the L X-ray series would be most important. In X-ray spectra of monazite samples these transitions are buried under the strong K X-ray peaks of the lanthanide elements. They should, however, become visible after chemical enrichment of the superheavy elements[2]. Such a treatment would also solve background problems due to a γ ray at 27.2 keV from the ^{140}Ce(p, n)^{140}Pr nuclear reaction[3–6] and to K X rays of ordinary trace elements[7]. Although our search remained negative we wish to report the results because our approach differs from that followed by others[3,8–14] in experiments which also remained negative or did not provide compelling evidence[6,15].

Due to a lack of monazite inclusions surrounded by giant haloes we have used bulk monazites as there is no obvious geochemical reason that superheavy elements should not occur in such material if they exist in inclusions although their concentration may be considerably lower First, we inspected a variety of samples, 18 monazites (from Madagascar, Nigeria, South Africa, Malaysia, Singapore, Thailand, Australia, Brasilia, Uruguay and Norway) and 2 bastnasites (from Madagascar), for superheavy element X rays by photon-induced X-ray fluorescence using the 60-keV γ ray of ^{241}Am for excitation When low-energy γ radiation rather than protons are used for fluorescence excitation, interferences by nuclear reactions can be ruled out 35-mg samples of the material were mounted on filter paper and exposed to an annular ^{241}Am source which was covered with a 0 05-mm thick rhodium foil to absorb the interfering weak 26 35-keV γ radiation of ^{241}Am The X-ray spectra were measured with a Si(Li) detector having a resolution of 0 175 keV at 5 9 keV Two spectra were taken per sample, one for 20 min to determine main constituents, and one for 10 h for trace elements In the longer exposures, spectra with good statistical quality (about 10^4 counts per channel) were obtained in the most interesting region between 23 and 29 keV

Table 1 Concentrations of some trace elements in monazites

Locality	Type of sample	In	Sn	Sb	Te
		Concentration† (p p m)			
Madagascar	Bulk	0 8	10	0 2	0 15
	Crystal*	0 3	1 0	1 2	≤0 03
Nigeria	Sand	0 9	350	0 3	0 1
Australia	Sand	0 7	12	1 2	0 9

* Collected[3] in the region of Ambatofotsikely where giant haloes occur in the surrounding biotite

† Average values, due to inhomogenities individual values can vary up to one order of magnitude

No evidence for a peak at 27 25 keV energy was found in any of the samples This corresponds to a concentration limit of ~30 p p m for element 126 in bulk monazites and also in bastnasites, a mineral which strongly enriches natural plutonium[16], the most likely chemical homologue of element 126 (ref 17 and B Fricke, personal communication) This is about the same limit (6–45 p p m) placed by the proton-induced X-ray fluorescence analysis of monazite inclusions[6] Clearly present in many samples was the tin $K_{\alpha1}$ X-ray peak, some samples also showed X-ray peaks of indium and antimony

To improve the detection limit significantly, six different samples of monazites (from Madagascar, Nigeria, Malaysia and Australia) and one sample of bastnasite (from Madagascar) were selected for chemical treatment As the chemical behaviour of elements belonging to the superactinide series with 6f–5g electrons cannot be predicted with certainty[17] the procedure applied takes into account that the elements searched for may follow the lanthanides, thorium, uranium or plutonium, even the unlikely case that they form insoluble sulphides or accompany alkaline earth and alkali elements has not been excluded *a priori* The procedure adopted is shown in Fig 1 It essentially consists of the following steps (1) 20-g samples were dissolved in hot concentrated sulphuric acid The insoluble part was fused with sodium peroxide, dissolved in water and hydrochloric acid and combined with the main sample (2) Sulphides were co-precipitated with copper sulphide from acid solution, this fraction should also contain superheavy elements around element 114, eka lead (3) By a hydroxide precipitation the bulk of the cationic elements was separated from alkaline earth and alkali elements (4) After dissolution of the hydroxides uranium- and plutonium-like elements were oxidized and lanthanides and thorium were precipitated as fluorides (5) The plutonium fraction was precipitated by ammonia in presence of carbonate ions which kept uranium in solution from where it was finally collected (6) The fluorides were dissolved in hydrochloric acid plus boric acid and transferred by a hydroxide precipitation into a strong nitric acid solution from which thorium iodate was precipitated, the lanthanides were collected in the filtrate Bastnasites were processed by a very similar procedure Samples of the different fractions obtained were

0028-0836/82/190136—03$01 00

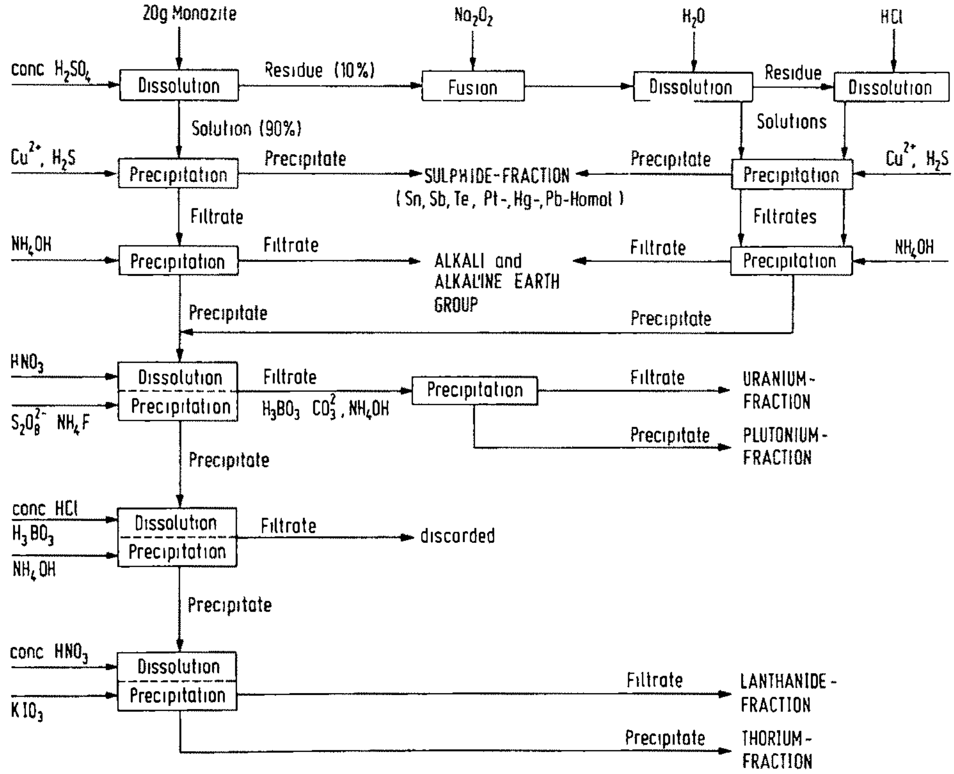


Fig. 1 Flow diagram of the chemical procedure applied in the search for superheavy elements in monazites and bastnasites

mounted for X-ray fluorescence analysis as described for bulk samples

As an example for the spectra obtained after chemical enrichment, Fig 2 shows the spectrum of a plutonium fraction No peaks can be seen at the positions of the $L_{\alpha 1}$ and $L_{\beta 2}$ X-ray transitions of element 126 From such spectra, upper limits of about 5–20 p p b (at 95% confidence level) are estimated for plutonium-like superheavy elements in the monazites investigated Peaks of main constituents are also absent, hence decontamination factors of $>10^7$ have been achieved in the chemical enrichment Similar concentration limits are estimated for the unlikely case that elements around $Z=126$ would be collected in the sulphide fraction For the other fractions, the X-ray fluorescence measurements are less sensitive because only a part of the total fraction could be mounted We estimate upper concentration limits between 1 and 30 p p m for the different fractions These limits hold for both monazites and bastnasites and should be considered as crude estimates

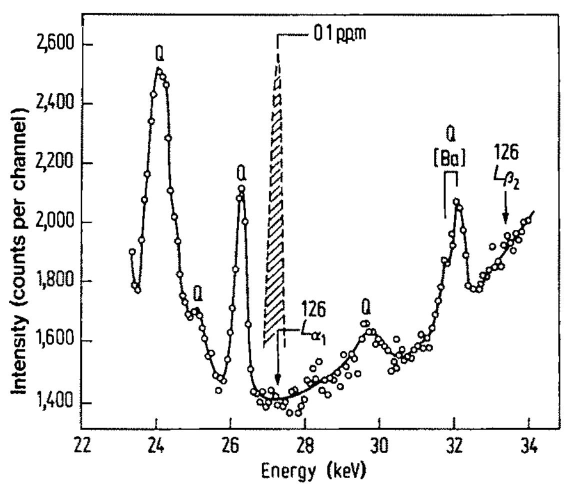


Fig. 2 Photon-induced X-ray fluoresence spectrum of a plutonium fraction isolated from a Madagascan monazite The hatched peak at the position of the $L_{\alpha 1}$ line of element 126 would result for a concentration of 0 1 p p m in the original bulk monazite sample Peaks denoted by Q belong to the background stemming from the ^{241}Am photon source and from trace elements (barium) in the filter paper used for sample mounting

The concentrations of the elements indium, tin, antimony, and tellurium which have K X-ray peaks in the region of interest were determined in some samples by X-ray fluorescence Tin and antimony were measured in the sulphide precipitate of the procedure Fig 1 As part of the tellurium(VI) is found in the hydroxide precipitate, this precipitate was dissolved in concentrated hydrochloric acid, and tellurium(VI) was reduced by boiling and then co-precipitated with copper sulphide Indium was determined in a separate sample after extraction from sulphuric acid into diethyldithiophosphoric acid and back extraction with strong hydrochloric acid Chemical yields were measured with radioactive tracers Residues in the dissolution of monazites were treated separately The results are given in Table 1 With the exception of tin the concentrations are too low to produce peaks in our spectra of unseparated monazites However, more than one order of magnitude higher concentrations were found for antimony and tellurium in X-ray spectra excited by protons[7]

In a selected number of fractions, mainly plutonium fractions, α particle spectra were recorded with semiconductor detectors No evidence for low-energy emitters reported long ago[18] nor for high-energy α particles[19] postulated as an explanation for the giant haloes[1] was found, in accordance with other studies[20,21] Inspection of some fractions by mass spectrometry also did not reveal any anomalies

We thank H Ahrens, F Begemann, W Bruchle, A -K Kaffrell, N Kaffrell, J V Kratz, H Menke, W Stegmann and M Weber for help, and the Treibacher Werke, K Chayawattanangkur, and D Schwalm for supplying monazite samples Financial support by the Studienstiftung des Deutschen Volkes (R S), the Bundesministerium fur Forschung und Technologie and by the Gesellschaft fur Schwerionenforschung Darmstadt is gratefully acknowledged

Received 16 November 1981, accepted 16 March 1982

1 Gentry, R V *et al* *Phys Rev Lett* **37**, 11 (1976)
2 Heimann, R *et al* *Ges fur Schwerionenforschung Darmstadt A Rep*, 73 (1976)
3 Bosch, F *et al* *Z Phys* **A280**, 39 (1977)
4 Fox, J D *et al* *Phys Rev Lett* **37**, 629 (1976)
5 Fletcher, N R *Phys Rev Lett* **38**, 479 (1977)
6 Cahill, T A *et al* *Phys Rev* **C17**, 1183 (1978)

7 Wolfli, W et al J Phys G3, L33 (1977)
8 Stephan, C et al Phys Rev Lett 37, 1534 (1976)
9 Ketelle, B H et al Phys Rev Lett 37, 1734 (1976)
10 Stoughton, R W et al J inorg nucl Chem 41, 1655 (1979)
11 Sparks, C J et al Phys Rev Lett 38, 205, 617 (1977), 40, 507 (1978)
12 Jelley, N A et al Nature 265, 35 (1977)
13 Middleton, R et al Phys Rev C16, 477 (1977)
14 Annegarn, H J et al Phys Rev C16, 379 (1977)
15 Cookson J A et al in Proc int Symp on Superheavy Elements (ed Lodhi, M A K) 164 (Pergamon, New York, 1978)
16 Hoffman D C, Lawrence, F O Mewherter, J L & Rourke, F M Nature 234, 132 (1971)
17 Fricke B Structure Bonding 21, 89 (1975)
18 Gysae B Z Naturforschung 5a, 530 (1950)
19 Chevallier, A, Chevallier, J, Pape, A & Debeauvais, M J Phys 38, L331 (1977)
20 Hirdes, D et al J Phys 40, L97 (1979)
21 Fireman, E L in Proc int Symp on Superheavy Elements (ed Lodhi, M A K) 172 (Pergamon, New York, 1978)

Environment of Ca^{2+} ions in aqueous solvent

N. A. Hewish*, G. W. Neilson† & J. E. Enderby

H H Wills Physics Laboratory, University of Bristol, Royal Fort, Tyndall Avenue, Bristol BS8 1TL, UK

The ion Ca^{2+} is common in nature and has an important role in many biochemical processes[1]. However, the way in which this ion behaves in aqueous solution is the subject of much controversy. There are two structural parameters which are necessary for a fuller understanding of Ca^{2+} in solution: the number of water molecules to which Ca^{2+} coordinates in solution, $\bar{n}$, and the configuration which a water molecule adopts relative to it (defined by ϕ, the angle between the plane of the water molecule and the Ca–O axis and r_{CaO}, the Ca–O separation). The technique of neutron diffraction in conjunction with isotopic substitution has been successful in giving unambiguous information regarding $\bar{n}$, ϕ and r for a variety of aqueous solutions[2]. Studies have been carried out on Li^+, Ni^{2+} and Cl^- at several concentrations[3] and Ca^{2+} at one concentration[4]. We report here the results of a study of the structure of the Ca^{2+}–D_2O conformation as a function of concentration.

Investigations have been carried out on several calcium halide and calcium nitrate water systems using X-ray diffraction[5–9], and the results have been used to obtain r_{CaO} (ref 2) However, these experiments have not produced information about the concentration dependence of $\bar{n}$ and ϕ, the central problem which we consider here The isotopic substitution made was $^{nat}Ca \rightarrow {}^{44}Ca$ which changes the neutron scattering length b from 4 90 fm to 1 80 fm

Neutron diffraction data were gathered at room temperature on the D4 diffractometer of the ILL, Grenoble for two solutions of $CaCl_2$ in heavy water identical in all respects except for the isotopic state of the calcium (Table 1) Total structure factors, $F(k)$, were derived from the data after correction for multiple scattering and absorption, and were normalized by reference to the scattering from a vanadium standard[2]

The first-order difference $\Delta(k)$ between the two $F(k)$s for a given pair of solutions was derived $\Delta(k)$ is a linear combination of partial structure factors $S_{ij}(k)$ and has the form[2]

$$\Delta(k) = A[S_{CaO}(k)-1] + B[S_{CaD}(k)-1] + C[S_{CaCl}(k)-1] + D[S_{CaCa}(k)-1]$$

where c_i = atomic fraction of species i, b_i = coherent scattering length of species i, $A = 2c_{Ca}c_O\,\Delta b_{Ca} b_O$, $B = 2c_{Ca}c_D\,\Delta b_{Ca} b_D$, $C = 2c_{Ca}c_{Cl}\,\Delta b_{Ca} b_{Cl}$, $D = c_{Ca}^2\,\Delta b_{Ca}^2$, $\Delta b_{Ca} = {}^{nat}b_{Ca} - {}^{44}b_{Ca}$, $\Delta b_{Ca}^2 = {}^{nat}b_{Ca}^2 - {}^{44}b_{Ca}^2$

Fourier transformation of this function leads to the weighted radial distribution function, $\bar{G}(r)$ given by

$$\bar{G}(r) = A(g_{CaO}(r)-1) + B(g_{CaD}(r)-1) + C(g_{CaCl}(r)-1) + D(g_{CaCa}(r)-1)$$

where

$$g_{ij}(r) - 1 = \frac{1}{2\pi^2 \rho r}\int (S_{ij}(k)-1)\sin kr\, k\, dk$$

Here $g_{ij}(r)$ is the correlation function relating to the distribution of j-type particles about i-type particles and ρ is the total number density of the solution $\bar{G}_{Ca}(r)$ is shown in Fig 1, the characteristic twin peaks close to 2 4 Å and 3 0 Å are associated with the oxygen and deuterium atoms respectively of the water molecules forming the first coordination shell around the Ca^{2+} ion Within this first shell, the contribution from the ion–ion

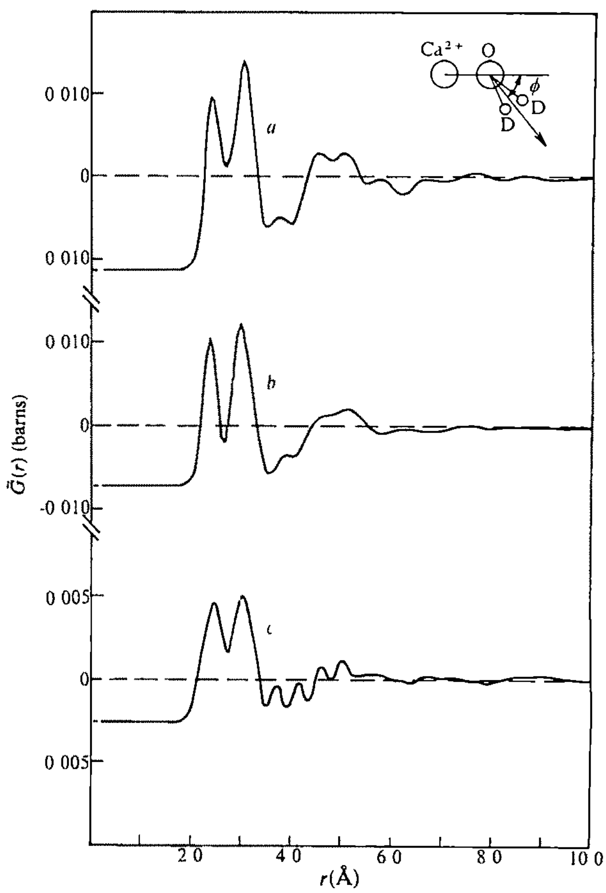


Fig 1 The weighted distribution $\bar{G}(r)$ for three $CaCl_2$ solutions in D_2O (a) 4 5 molal (b) 2 8 molal (c) 1 molal The ion–water geometry is shown at the top right

distribution functions $g_{CaCl}(r)$ and $g_{CaCa}(r)$ are negligible because the coefficients C and D are small relative to A or B (Table 1) The number of atoms, $\bar{n}_i$, to which the areas of any given peak corresponds is calculated using

$$\bar{n}_i = 4\pi c_i \rho \int_{r_1}^{r_2} r^2 g_{Cai}(r)\, dr$$

where r_1 and r_2 are the positions of the minima adjacent to the peak on either side

The above identification of the twin peaks is confirmed by the calculation of $\bar{n}_O$ and $\bar{n}_D$, in which it is found that $2\bar{n}_O = \bar{n}_D$, within experimental error The separation of the two peaks may be used to calculate ϕ assuming a geometry in which r_{OD}

* Present address Standard Telephones Ltd London Road, Harlow, Essex CM17 9NA, UK
† To whom correspondence should be addressed

0028-0836/82/190138—02$01 00

Table 1 Scattering lengths and sample parameters for $CaCl_2$ heavy water solution

Molality	c	Isotopes	Scattering lengths (fm)	A	B	C	D
				($\times 10^{-3}$ barns)*			
1.0	0.0065	^{nat}Ca	4.90	0.77	1.77	0.05	0.01
		^{44}Ca	1.80				
2.8	0.0177	^{nat}Ca	4.90	2.03	4.66	0.38	0.07
		^{44}Ca	1.80				
4.5	0.0275	^{nat}Ca	4.90	3.02	6.95	0.90	0.16
		^{44}Ca	1.80				

* 1 barn $= 10^{-24}$ cm^2

Table 2 Ca^{2+}–water coordination

c(molal)	r_{CaO}(Å)	r_{CaD}(Å)	ϕ	$\bar{n}$
1	2.46±0.03	3.07±0.03	38°±9	10.0±0.6
2.8	2.39±0.02	3.02±0.03	34°±9	7.2±0.2
4.5	2.41±0.03	3.04±0.03	34°±9	6.4±0.3

is 1Å, and the angle DÔD is 105°. The results of the two latest experiments are summarized in Table 2 and Fig. 1, where we include those derived from an earlier study of a 4.5 molal $CaCl_2$.

When we compare these results with those, at comparable concentrations, for a strongly hydrating divalent ion Ni^{2+} (ref. 10) we find the following points of similarity and difference: (1) r_{CaO}, like r_{NiO} does not, when allowance is made for experimental error vary significantly with concentration. In both cases the distances agree within experimental error with those derived from X-ray studies[5–9], (2) for both ions at concentrations in excess of 1 molal, ϕ is independent of ionic strength and is ~35° which is significantly different from both the 'dipole' configuration ($\phi = 0$) and the 'lone pair' configuration ($\phi = 55°$), (3) for Ni^{2+}, $\bar{n}$ is independent of concentration. For Ca^{2+}, by contrast, $\bar{n}$ is strongly concentration dependent, increasing from ~6 to ~10 as the molality decreases from 4.5 to 1, (4) the widths of the first peak for both $g_{CaO}(r)$ and $g_{CaD}(r)$ decrease with increase in concentration, this behaviour is opposite to that observed for $g_{NiO}(r)$ and $g_{NiD}(r)$.

We believe that points (3) and (4) are manifestations of the weakly hydrating nature of Ca^{2+} ions in solution. Unlike Ni^{2+} where our earlier experiments have shown the almost universal nature of the Ni^{2+}–H_2O conformation[11] the aqueous environment of Ca^{2+} will depend crucially on the ionic strength, the counter ion and the temperature. It follows that if the properties of Ca^{2+} in solution are to be understood properly a range of experiments of the type we have reported here must be undertaken in which the principal thermodynamic variables (concentration, pressure and temperature) are changed systematically.

N. A. H. thanks the SERC for a research studentship. We all thank the scientific and technical staff at the ILL, Grenoble, and particularly Dr Chieux and Dr Cummings, for valuable help with the experiments. The aqueous solution programme of research is financed by the SERC and we are grateful for their support.

Received 15 February, accepted 16 March 1982

1. Dick, D. A. T. in *Water and Aqueous Solutions* (ed. Horne, R. A.) Ch. 7 (Wiley Interscience, New York, 1972)
2. Enderby, J. E. & Neilson, G. W. *Water, a Comprehensive Treatise* Vol. 6 (ed. Franks, F.) 1 (Plenum, New York, 1979)
3. Enderby, J. E. & Neilson, G. W. *Rep. Progr. Phys.* **44**, 593 (1981)
4. Cummings, S., Enderby, J. E. & Howe, R. A. *J. Phys. C Solid State Phys.* **13**, 1 (1980)
5. van P. van Eck, C. L., Mendel, H. & Boog, W. *Disc. Faraday Soc.* **24**, 200 (1957)
6. Albright, J. N. *Chem. Phys.* **56**, 783 (1972)
7. Licheri, G., Piccaluga, G. & Pinna, G. *J. Chem. Phys.* **64**, 2437 (1976)
8. Bol, W., Gerrits, G. J. A. & van P. van Eck, C. L. *J. appl. Crystallogr.* **3**, 486 (1970)
9. Kollman, P. & Kuntz, I. D. *J. Am. chem. Soc.* **94**, 9236 (1972)
10. Neilson, G. W. & Enderby, J. E. *J. Phys. C* **11**, L625 (1978)
11. Newsome, J., Neilson, G. W., Enderby, J. E. & Sandström, M. *Chem. Phys. Lett.* **82**, 399 (1981)

Evolution of passive continental margins and initiation of subduction zones

S. A. P. L. Cloetingh, M. J. R. Wortel & N. J. Vlaar

Vening Meinesz Laboratory, Instituut voor Aardwetenschappen, University of Utrecht, Budapestlaan, 4, 3584 CD Utrecht, The Netherlands

Although the initiation of subduction is a key element in plate tectonic schemes for evolution of lithospheric plates, the underlying mechanisms are not well understood. Plate rupture is an important aspect of the process of creating a new subduction zone, as stresses of the order of kilobars are required to fracture oceanic lithosphere[1]. Therefore initiation of subduction could take place preferentially at pre-existing weakness zones or in regions where the lithosphere is prestressed. As such, transform faults[2,3] and passive margins[4,5] where the lithosphere is downflexed under the influence of sediment loading have been suggested. From a model study of passive margin evolution we found that ageing of passive margins alone does not make them more suitable sites for initiation of subduction. However, extensive sediment loading on young lithosphere might be an effective mechanism for closure of small ocean basins.

The state of stress at a passive margin is determined by its local geometrical and rheological lithospheric properties and by the system of forces acting on the lithosphere. Of these features the thickness of the oceanic lithosphere[6], its rheological stratification[7], the push exerted by the elevation of the oceanic ridge[8,9] and the forces associated with the (negative) buoyancy of the lithosphere[10] are a function of the age of the oceanic lithosphere. The sediment loading capacity of oceanic lithosphere increases with age, through its continued cooling and densification. One might therefore expect a coupling between the height of the sedimentary column deposited at the passive margin and the age-dependent thermal subsidence of the underlying oceanic lithosphere[11]. In previous studies of the state of stress at passive margins[12–14] the possible implications of age-dependent properties of the oceanic lithosphere were not taken into account. We have studied the interrelations between age-dependent forces, geometry and rheology, to decipher their net effect on the state of stress at passive margins. We have constructed finite element models for a passive continental margin in different stages of evolution at ages of 20, 30, 60, 100 and 200 Myr. For all models a half spreading rate of 1 cm yr^{-1} is taken, characteristic of oceanic lithosphere without downgoing slabs attached to it[15]. The model features are summarized in Fig. 1. As a model for sedimentary loading we adopt triangular wedges at the continental shelf and rise. As our reference we assume that the maximum thickness of the sediments at the margin corresponds with the thickness that results if the sedimentation has been keeping up with the subsidence of a boundary layer model of the cooling oceanic lithosphere[11,16]. This implies for the maximum height of the sedimentary triangular wedge an increase from 3.6 km at 20 Myr to 9.4 km at 200 Myr, following roughly a square root of age relation. Observational data on thicknesses of post-rift sedimentary sequences are now available from geophysical surveys carried out on passive margins during the past few years[17]. From these data it follows that the reference model is representative of most of the sediment loading histories and resulting thicknesses observed at passive margins. The huge sediment accumulations at deltas, however, clearly exceed the thicknesses as given by the reference model. Therefore, a second class of models was constructed in which the full loading capacity of oceanic lithosphere was taken up by sediments with the sedimentary thicknesses ranging from 10 to 15.8 km. The width of the

sedimentary wedges varies between 150 km for very young margins to 250 km for mature margins, based on data compilations[17] As the sediments replace water we consider only excess densities ($\Delta\rho = 2\ 4 - 1 = 1\ 4\ \mathrm{g\ cm^{-3}}$)

Studies on seismicity and flexure of oceanic lithosphere bending at trenches[18 19] and analysis of flexure of the oceanic lithosphere under the influence of seamount loading[20] have demonstrated that the lithosphere is capable of supporting stresses of the order of some kilobars on geological time scales The experimental work of Goetze and co-workers[21 22] shows that the strength of the lithosphere first increases with depth and then decreases rapidly to a level of a few hundred bars in the lower part of the plate, where temperature effects become dominant and the maximum stress achievable is limited by thermally activated ductile flow Deriving temperature profiles from Crough's[23] model for the oceanic lithosphere, which model combines the merits of the boundary layer model and the plate model, we have constructed lithospheric strength envelopes for a wide range of ages These are based on Goetze's ductile flow laws in olivine for power-law and Dorn-creep rheologies, assuming a strain-rate of $e = 10^{-18}\ \mathrm{s^{-1}}$ For the sediments we assume a zero strength A similar approach was followed by Bodine and co-workers[24] Yield envelopes for lithospheric ages of 30 and 100 Myr are given in Fig 2 which shows that both the thickness H of the 'mechanical plate' (given by the depth below which the strength of the plate is <0 5 kbar) and the maximum strength increase strongly with age The flexure of the lithosphere is counteracted by isostasy Isostatic forces proportional to the deflection due to loading are included in the model Of the plate tectonic forces implemented in the models those associated with the ridge push and the negative buoyancy of the oceanic lithosphere are calculated on the basis of Oxburgh and Parmentier's[25] model for the formation of oceanic crust and Crough's[23] model for the thermal evolution of oceanic lithosphere We ignore drag at the base of the lithosphere Zero horizontal displacements are prescribed for the right-hand boundary of the model to simulate a ridge push transmitted through the continent from an adjacent oceanic plate

The deformation of the lithosphere at passive margins and the resulting stress field (order of magnitude of a few kilobars) are dominated by sediment loading[26] The contribution of the plate tectonic forces to the stress field is an order of magnitude smaller Differential stresses are largest at the points of maximum flexure These are located under the rise in the oceanic plate close to the transition of oceanic and rift-stage lithosphere Figure 2a, b shows the stress maxima for 100 and 30 Myr for the reference load The lowermost part of the mechanical plate is in yield due to the tensile stresses developed at the base The main part of the plate remains in the elastic state The effect of the rheological stratification of the lithosphere is twofold stress relaxation in the lowermost part and stress concentration in the mechanically stronger upper part This effect is particularly important if the full loading capacity of the lithosphere is taken up by sediments This can even result in complete failure of the lithosphere as demonstrated in Fig 2c

Figure 2 also shows that the state of stress depends on the age of the margin To illustrate this dependence more clearly we have plotted in Fig 3a the maximum differential (tensional)

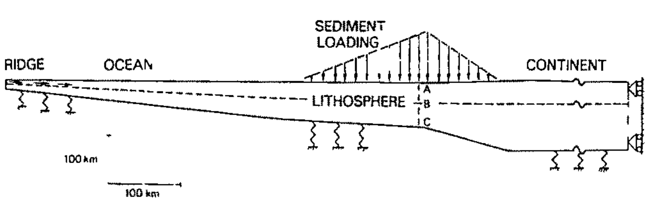


Fig 1 Model features geometry, rheology, system of forces and boundary conditions Mechanical thickness indicated by broken horizontal line Young's modulus $E = 7 \times 10^{10}\ \mathrm{N\ m^{-2}}$ and Poisson's ratio $\nu = 0\ 25$

stresses as a function of age The age dependence is strongest for ages below 100 Myr From 30 to 100 Myr—an interval in which the sedimentary loading, the mechanical thickness and the strength increase—the differential stress maxima increase with age From 100 to 200 Myr the mechanical thickness and the strength of the plate show only a small increase For ages

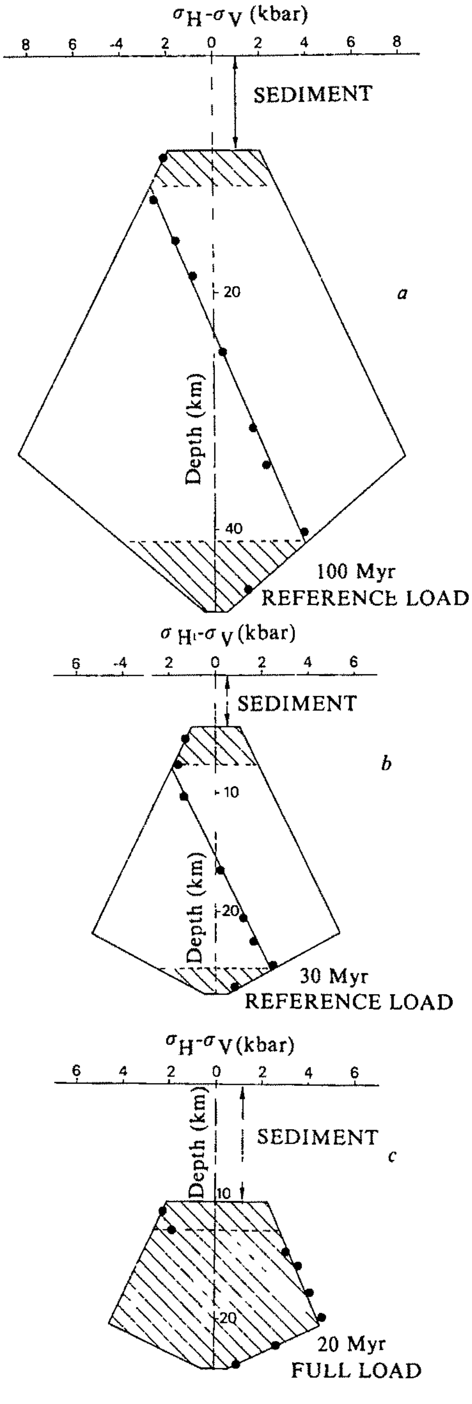


Fig. 2 Strength envelope and results of model calculations (●) at cross-section through the upper part of the lithosphere at the point of maximum flexure (AB of cross-section ABC in Fig 1) The line inside strength envelope connecting the solid dots is the stress distribution Differential stresses ($\sigma_H - \sigma_V$) are plotted versus depth Sign convention for the stresses tension positive, compression negative Zero-strength has been assumed for the sediments Hatched areas in the upper and lower part of the mechanical layer denote failure by brittle fracture and ductile flow respectively *a*, *b*, The results for the reference model of sediment loading for ages of 100 and 30 Myr *c*, The results for the full load model for 20 Myr The horizontal dashed line indicates the neutral surface just before complete failure of 20-Myr old lithosphere takes place

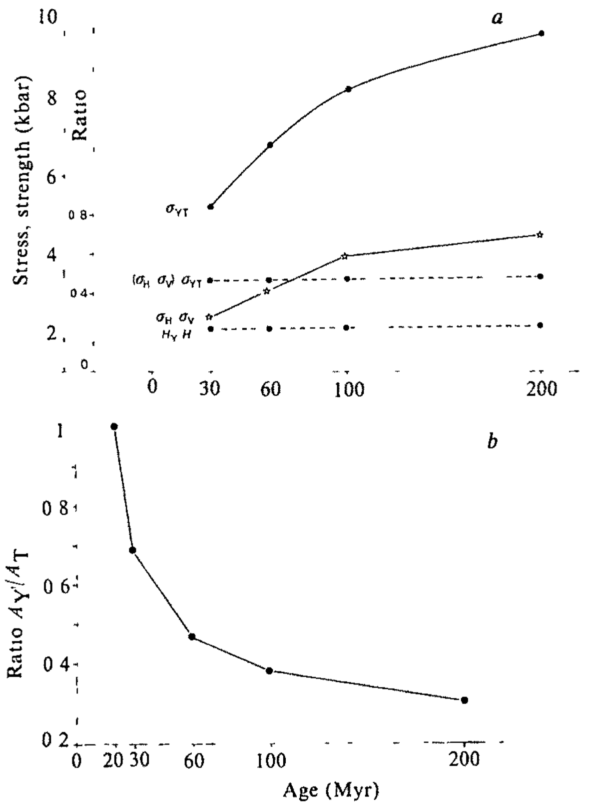


Fig 3 *a*, Model results for reference load maximum differential stresses ($\sigma_H - \sigma_V$), tensile strength σ_{YT}, their ratio and the relative thickness of the mechanical layer in failure (H_Y/H) plotted as a function of lithospheric age *b*, Results full load model ratio of A_Y (the hatched area inside the strength envelope, see Fig 2) and A_T (total area of the envelope) as a function of lithospheric age For ages below 20 Myr complete lithospheric failure is induced

above 100 Myr the increase in sediment loading according to our reference model results in only a minor increase of the stresses Interesting quantities are the ratio of the maximum stress generated and the maximum strength, and the ratio of A_Y, corresponding with the hatched area (inside the strength envelope (Fig 2)) and the total area A_T of the envelope For the reference model of sediment loading these quantities prove to be essentially independent of age because load, strength and plate thickness all exhibit the same square root of age behaviour due to the thermal evolution of the plate The situation changes drastically if the full loading capacity is taken up The surplus load of sediments added to the reference load is most effective in creating high stresses when deposited on a young margin Figure 3*b* and results of numerical calculations made for ages below 20 Myr (not shown here) demonstrate that full loading on a young passive margin (age below 20 Myr) leads to complete failure of the lithosphere Owing to the stabilizing affect of the density changes accompanying the formation of oceanic crust[25] oceanic lithosphere is gravitationally stable for ages <20–40 Myr (ref 16) As a result of its further cooling and thermal contraction the lithosphere becomes unstable for ages above 20–40 Myr (refs 16, 25) Taking into account only the lithospheric instability it was reasonable to expect that chances for initiation of subduction of oceanic lithosphere would increase with the age of the margin[10] Our work shows that if after a short evolution of the plate, subduction has not yet started, continued ageing of the passive margin alone does not result in conditions more favourable for plate rupture and initiation of subduction

Although most of the present deltas of the world are deposited on old oceanic lithosphere[27] and very few examples of modern thick sedimentary cones on young lithosphere are known, extensive sedimentary loading might have been an effective mechanism for closure of small ocean basins in geological history Closure of small oceanic basins has an important role in the process of mountain building In this context, note that in a comprehensive account of the evolution of the central Alps, Frisch[28] presents evidence for closure of oceanic basins within the first 100 Myr after opening

For older margins in general, however, it seems that the stresses generated are insufficient to induce lithospheric failure and initiation of subduction Therefore, pre-existing weakness zones in oceanic lithosphere might be more suitable sites for initiation of subduction than passive margins This view is consistent with the results of a survey of recently initiated subduction zones in the Pacific[29] showing that zones initiated in the Neogene are either at the sites of transform faults or rejuvenated pre-existing subduction zones Many of the present circum-Pacific zones are in fact the successors of subduction zones already present in the configuration before the breakup of Gondwanaland Further back in geological time different plate configurations and a different thermal regime might have provided conditions more suitable for initiation of subduction As an extreme example one might consider the Archaean where, due to the much steeper temperature profiles[30], the lithosphere must have been considerably weaker than nowadays In such a situation plate rupture requires a considerable lower stress level This suggestion is corroborated by geological data on Precambrian orogenic belts[31], showing extensive activation of passive margins bordering small oceanic basins

In the present lithospheric system special circumstances may arise in the process of global plate reorganization, when plate tectonic forces may be concentrated locally This can be important in an oceanic plate attached to a subduction zone where the pull acting on the subducting slab can be concentrated to a high level (order of magnitude of several kilobars)[32] In particular Wortel and Cloetingh[32] have shown that lateral variations in the age of the slab descending in a subduction zone might provide a mechanism for fragmentation of oceanic plates and the formation of spreading centres In an opening oceanic basin, however, the only plate tectonic force to be concentrated is the ridge push, inducing stresses with an order of magnitude of only a few hundred bars Although the concentration factor might be quite high[33] the level of the concentrated stresses due to ridge push is by no means comparable with the stresses resulting from sediment loading Therefore, plate reorganization might take place predominantly by the formation of new spreading ridges In such a process new subduction zones might be subsequently created at the sites of transform faults already present in the plate, when the new spreading direction has a component perpendicular to the direction of the transform fault

We thank Gerald Wisse (Delft University of Technology) for support with the finite element calculations, and Joel S Watkins and Rob van der Voo for critically reading an earlier version of the manuscript

Received 19 January accepted 8 March 1982

1 Kirby S H *J geophys Res* **85**, 6353 (1980)
2 Uyeda, S & Ben-Avraham, Z *Nature* **240**, 176 (1972)
3 Dewey, J F *Am J Sci* **275A**, 260 (1975)
4 Dietz, R S *J Geol* **71**, 314 (1963)
5 Dewey, J F *Earth planet Sci Lett* **6**, 189 (1969)
6 Parsons, B & Sclater, J G *J geophys Res* **82**, 803 (1977)
7 Caldwell, J G & Turcotte, D L *J geophys Res* **84**, 7572 (1979)
8 Richter, F & McKenzie, D *J Geophys* **44**, 441 (1978)
9 England, P & Wortel R *Earth planet Sci Lett* **47**, 403 (1980)
10 Vlaar, N J & Wortel, M J R *Tectonophysics* **32**, 331 (1976)
11 Turcotte D L & Ahern J L *J geophys Res* **82**, 3762 (1977)
12 Walcott, R I *Bull geol Soc Am* **83**, 1845 (1972)
13 Bott, M H P & Dean, D S *Nature phys Sci* **235**, 23 (1972)
14 Turcotte, D L, Ahern, J L & Bird, J M *Tectonophysics* **42**, 1 (1977)
15 Forsyth, D W & Uyeda, S *Geophys J R astr Soc* **43**, 163 (1975)
16 Wortel M J R Thesis Utrecht Univ (1980)
17 Watkins J S, Montadert, L & Dickerson, P W *Am Ass petrol Geol Mem* **29** (1979)
18 Chapple, W M & Forsyth D W *J geophys Res* **84**, 6729 (1979)
19 McAdoo, D C, Caldwell, J G & Turcotte, D L *Geophys J R astr Soc* **54**, 11 (1978)
20 Watts, A B Bodine, J H & Steckler, M S *J geophys Res* **85**, 6369 (1980)
21 Goetze, C *Phil Trans R Soc* **A288**, 99 (1978)
22 Goetze, C & Evans, B *Geophys J R astr Soc* **59**, 463 (1979)
23 Crough, S T *Nature* **256**, 388 (1975)
24 Bodine, J H, Steckler, M S & Watts, A B *J geophys Res* **86**, 3695 (1981)
25 Oxburgh, E R & Parmentier, E M *J geol Soc Lond* **133**, 343 (1977)
26 Cloetingh, S A P L Wortel, M J R & Vlaar, N J *Am Ass petrol Geol Mem* (in the press)

27 Kinsman, D J J in *Petroleum and Global Tectonics* (eds Fischer, A G & Judson, S) 83 (Princeton University Press, 1975)
28 Frisch, W *Tectonophysics* **60**, 121 (1979)
29 Karig, D Preprint, Cornell Univ
30 Sharpe, H N & Peltier, W R *Geophys J R astr Soc* **59**, 171 (1979)
31 Kroner A in *Precambrium Plate Tectonics* (ed Kroner, A) 56 (Elsevier, Amsterdam, 1981)
32 Wortel, R & Cloetingh, S *Geology* **9**, 425 (1981)
33 Kato, T , Shimazaki, K & Yamashina, K *Geophys J R astr Soc* **60** 377 (1980)

Geological evidence against the Shyok palaeo-suture, Ladakh Himalaya

Hakim Rai

Wadia Institute of Himalayan Geology, Dehra Dun-248001, India

Evolution of the Ladakh Himalaya is related to the Cretaceous–Tertiary subduction of the Indian oceanic plate below the southern margin of the Eurasian plate along the Indus suture zone[1–11]. The plate tectonic reconstruction suggests two stages of subduction due to the presence of a pre-Cretaceous suture along the Shyok valley[12,13]. This hypothesis is based on there being ophiolite belts in the Indus and Shyok valleys. The Indus suture zone is well established but there have been insufficient data on the Shyok suture to confirm the existence of the palaeo-suture. I have studied the Saltoro hills, in the western part of the Shyok valley, and report here the salient geological features. With a few exceptions, the igneous and sedimentary rocks indicate that the Indus and Shyok sutures had the same Cretaceous–Oligocene history. However, the volcanic eruption during the deposition of post-Oligocene molasse in the Saltoro hills of the Shyok valley is the youngest igneous activity in the whole of the Ladakh Himalaya. Hence the geological evidence described here does not accord with the theory of palaeo-suture.

Since the formulation of the plate tectonic theory the occurrence of the Indus ophiolites has been taken as evidence for subduction in that area It has been suggested that the Himalayan mountain chain came into existence due to the subduction of the Indian plate under the Asian plate during the Cretaceous–Tertiary period[1–13] The problem attained a new dimension when an earlier subduction was proposed along the Shyok valley[12,13] to the north of the Indus suture zone (Fig 1) Previous work on the Shyok area has been unsystematic[14–19] Note that rock masses, other than the igneous, occur as thrust slices and the rock units do not maintain lithological or structural continuity in this region Recent workers[16–19] researching in limited and localized areas, have further confused the geology of this area

The eastern part of the Saltoro hills comprises igneous and sedimentary rocks ranging from Cretaceous to post-Oligocene Andesites and basalts (the Shyok volcanics) form the oldest igneous rocks erupted before the deposition of flysch Pyroxenite and hornblende gabbro have intruded into basalt–andesite association along the northern part of the Saltoro hills and in the lower Shyok valley The assemblage of basic rocks along with metasedimentary rocks of pre-Cretaceous age—the southern limit of the Karakoram sediments[15]—has again been intruded by granitoids, causing thermal metamorphism and migmatization

Major igneous activity ended in the region with the emplacement of Tertiary batholithic granitoids of the Ladakh and Karakoram[20] Sandwiched between these igneous rocks are two distinct sedimentary belts of flysch and molasse Both these belts have thrust contacts with each other as well as with the igneous rocks and occur as detached masses The flysch forms a marginal part of the eastern Saltoro hills and separates the basic volcanics from the molasse It is dominated by shale and limestone with occasional sandstone or quartzite The rocks are tightly folded near the thrust contact and are generally dipping due north The limestone, particularly in the lower part of the flysch, is highly fossiliferous and has yielded the following fossils of Upper Cretaceous–Eocene age *Fasciolites oblonga* (d'Orbigny), *Nummulites* sp aff *N obtusus* Sowerby, *N atacicus* Leymerie, *Assilina* sp , *Dictyoconus* sp , *Lockhartia* sp , *Cyclammina* sp , *Cerriocava nilkanthi* Singh, '*Feddenia*' *garkhalensis* Mathur, *Lophosmilia* sp , *Girvanella* sp , *Lithophyllum* sp and charophyte gyrogonite Beside these characteristic fossils, the rocks have yielded turritellid gastropods, brachiopod shells and crinoid stems

A thin belt of molassic sediments, indicative of shallow continental conditions, contain rounded to sub-rounded pebbles and boulders of limestone, chert, basalt, andesite, rhyolite and granitoids The limestone boulders have yielded *Orbitolina* sp , *Dictyoconus* sp , *Nummulites* sp , ill-preserved gastropods (turritellides) and so on The granitic and rhyolitic clasts have been derived from the Tertiary batholiths (Ladakh and Karakoram) and Khardung acid volcanics (38 ± 2 Myr)[21] of the Shyok valley Hence the molasse is post-Oligocene Penecontemporaneous andesitic lava flows and subsequent intrusion of basic and acidic dykes are significant features of this molasse, which is folded and has thrust contacts with flysch in the south and basic volcanics (with associated granitoids) to the north Emplacement of discontinuous bodies of chromite-bearing serpentinites with black shales has taken place along the thrust contacts, especially along the northern border of the molasse

The Himalayan mountain chain has been considered to be a typical example of continent–continent collision[3] It is thought that the subduction of the north-moving Indian plate started along the Indus suture during the end of the Cretaceous and the actual collision of the continental parts of the two plates probably began in the late Eocene or early Oligocene[10,22] Recent studies by Frank *et al*[13] in the Shyok valley, to the north of the Indus suture zone, have also included all the volcanics (undifferentiated) with Shyok ophiolites and assigned an age older than the Indus ophiolites This led them to reconstruct a new plate tectonic model for the Himalayan mountain belt, proposing two stages of subduction separated in space and time Accordingly, the first subduction was initiated in the pre-Cretaceous time (exact time of subduction is not given) along the Shyok suture, which became inactive later The second subduction followed the Indus suture during the Upper Cretaceous–Tertiary period The later predominance of the Khardung acid volcanics younger than the Ladakh Batholith described from the Shyok valley[17,18] was opposed by Bhandari *et al*[16] and Thakur *et al*[19], who established the intrusion of batholithic units into these acid volcanics Hence, the Khardung volcanics are older than the Ladakh batholith They also separated the acid and basic volcanic rocks of this region as Khardung volcanics and the Shyok volcanics, respectively

I have found three distinct volcanic phases in the Shyok valley which differ in composition, age and the environment of eruption The oldest volcanic rocks, representing the Shyok ophiolites, are basic in composition and have suffered thermal metamorphism near the contacts with gabbro, pyroxenite, tonalite and granite This complex volcanic–plutonic association is partially overlain by the Upper Cretaceous–Eocene flysch and post-Oligocene molasse, a case similar to the Indus ophiolites Thus the Shyok valley marks the northern extremity of the Indus ophiolites as there is no evidence to indicate the separation of these two belts during the Cretaceous–Eocene Extensive occurrence of basic xenoliths within the Ladakh batholith suggests that the area between the Indus and Shyok valleys was occupied by the same ophiolitic rocks which are now associated with the Indus and Shyok ophiolites These rocks provided a thick cover for the batholithic intrusion Subsequent erosion has removed the cover rocks over the batholith Consequently, the Shyok ophiolites apparently form a separate belt that runs parallel to the Indus ophiolite belt A second volcanic eruption is documented by the Khardung volcanics having acidic composition and constituting a distinct belt along the northern margin of the Ladakh batholith between Hundar and Khardung villages (Fig 1), and extends eastwards beyond the area under discussion The batholithic units have intruded into these volcanics

0028–0836/82/190142—03$01 00

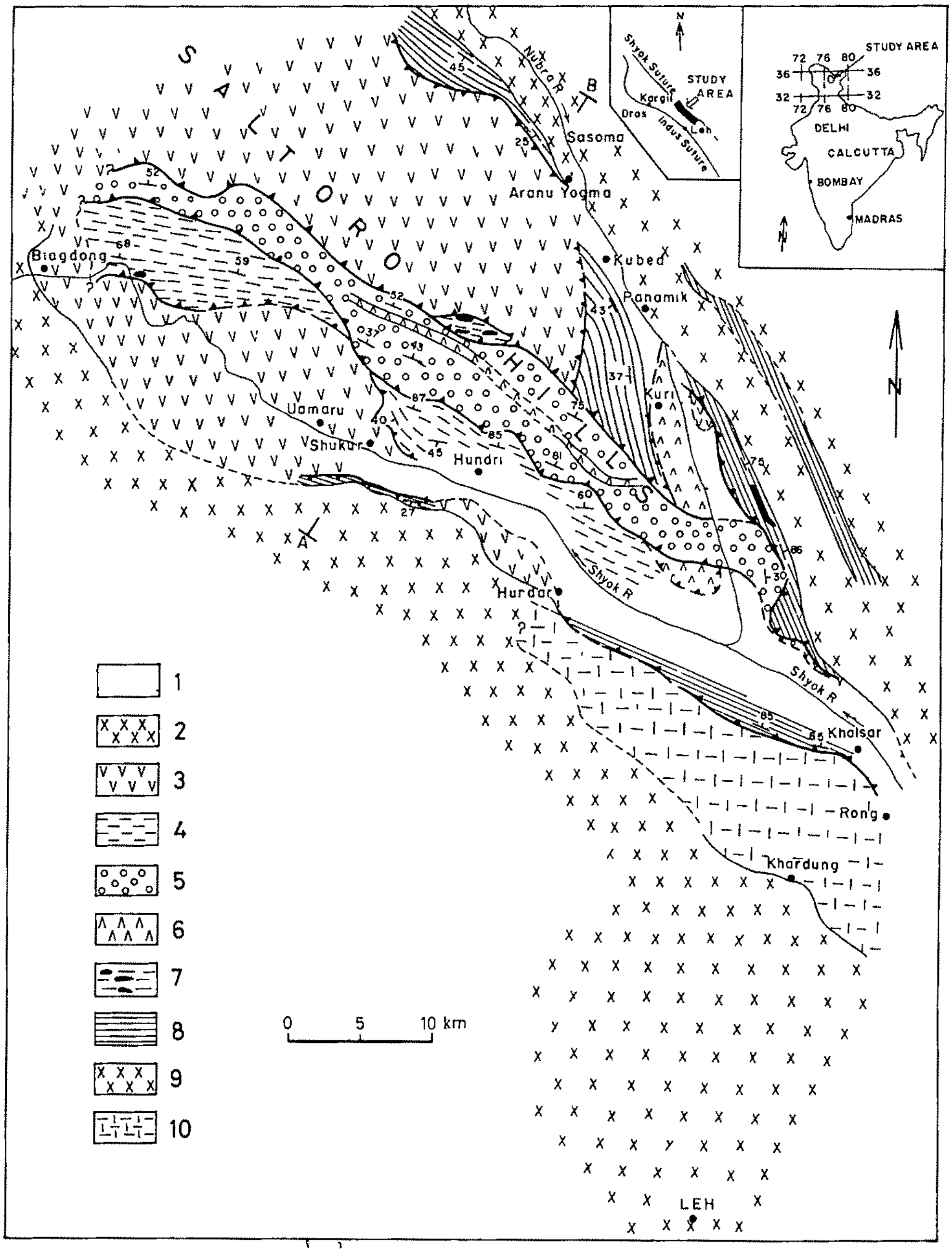


Fig 1 Geological map of the Shyok valley 1, river deposits, 2, granitoids of the Ladakh batholith, 3, Shyok volcanics with pyroxenite, 4, Cretaceous–Tertiary flysch, 5, molasse of Saltoro hills, 6, synsedimentary volcanics associated with molasse, 7, detached serpentinite bodies along with black shales, 8, Karakoram sediments (Permian?), 9, granitoids of the Karakoram batholith, 10, Khardung volcanics (acid volcanics)

at several places The association of lapilli tuffs and conglomeratic beds in the Khardung volcanics suggests that the rocks were erupted in shallow water to subaerial conditions Lapilli tuffs at the top of the main phase of this eruption are conformably overlain by lahars or mud flows having intermittent lava flows of rhyolitic composition The lahars are conformably overlain by coarse clastic, non-marine deposits, essentially composed of a sandstone–shale sequence with conglomerates

The third and the last phase of volcanic activity is preserved within the molasse of the Saltoro hills The reddish-brown volcanic rock is characterized by random distribution of euhedral plagioclase phenocrysts (An_{36-32}) in the fine groundmass Dykes of similar composition have intruded into the basic volcanics along the southern flank of Saltoro hills The molasse with which these andesitic volcanics are associated was deposited in a trough formed within basic volcanics which were intruded by the granitoids of the Ladakh batholith from the south and the Karakoram granitoids from the north

Because the molasse is rich in granitic material, probably derived from the Ladakh batholith (27.6 ± 0.6 Myr)[21], pebbles and boulders of acid volcanics (Khardung volcanics = 38 ± 2 Myr)[21] and limestone containing Palaeocene foraminifera (as in the flysch), it can be correlated with the Kargil molasse of Miocene to Pleistocene age exposed along the Indus suture zone In the absence of fossils no exact age can be assigned to the matrix of the molasse The extrusive and dyke rocks associated with molasse of the Saltoro hills are either Miocene or younger No such igneous activity has been observed in the molasse along the Indus suture zone This evidence suggests a prolonged igneous activity in the Shyok valley The volcanics and dykes of the Saltoro hills appear to be the youngest igneous rocks of the Ladakh Himalaya

It has been proposed that after subduction along the Shyok suture, a new phase of subduction commenced along the Indus suture zone during the Upper Cretaceous and continued up to the late Oligocene or early Miocene If the subduction along the Shyok suture stopped before the initiation of the new site of subduction along the Indus suture then the volcanic rocks in the Shyok valley should be much older than Upper Cretaceous It is

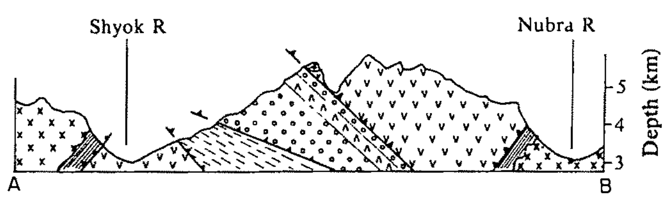


Fig. 2 The geological cross-section along A–B in Fig 1

evident from the above that the palaeo-suture idea[12,13] as proposed does not find any support lithologically, structurally or stratigraphically

A comparison of the present observations from the Shyok valley with the existing geological data from the Indus suture makes it possible to correlate the geological events of these two regions There has been widespread volcanic–plutonic igneous activity in the Ladakh Himalaya, extending from the Indus suture zone to the Karakoram during the initiation of the Himalayan orogeny The emplacement of basic and ultrabasic rocks took place from the Dras to Shyok valley on a regional scale The Ladakh batholith has intruded into this basic and ultrabasic rock association The existence of two parallel ophiolite belts is simply a result of batholithic intrusion and subsequent deep erosion, removing the cover rocks of the batholith which is rich in basic xenoliths Hence the Shyok ophiolites represent the northern extension of the Indus ophiolites emplaced during the Upper Cretaceous–Oligocene

I conclude that shallow marine conditions existed in the Shyok region at least up to the early Eocene Although the Indus suture zone became magmatically inactive during the end of the Oligocene, the Shyok valley represents still younger igneous activity Therefore, the present geological evidence does not favour the two-stage subduction with a palaeo-suture in the Shyok valley

I thank Professor I C Pande, N S Virdi and M I Bhat for comments, N S Mathur for identifying the fossil assemblage and the Director, Wadia Institute of Himalayan Geology, Dehra Dun for permission to publish this paper

Received 17 December 1981, accepted 16 March 1982

1 Dewey, J F & Bird, J M *J geophys Res* **75**, 2625–2647 (1970)
2 Powell, C McA & Conaghan, P J *Earth planet Sci Lett* **20**, 1–12 (1973)
3 Dewey, J F & Burke, K C A *J Geol* **81**, 683–692 (1973)
4 Crawford, A R *Geol Mag* **111**, 369–480 (1974)
5 Le Forte, P *Am J Sci* **275A**, 1–44 (1975)
6 Menke, H W & Jacob, K H *Bull seis Soc Am* **66**, 1695–1711 (1976)
7 Bird, P *J geophys Res* **83**, 4975–4987 (1978)
8 Armbruster, J, Seeber, L & Jacob, K H *J geophys Res* **83**, 269–282 (1978)
9 Klootwijk *et al Earth planet Sci Lett* **44**, 47–64 (1979)
10 Brookfield, M E & Reynolds, P H *Earth planet Sci Lett* **55**, 157–162 (1981)
11 Mitchell, A H G *J geol Soc Lond* **138**, 109–122 (1981)
12 Gansser, A *Coll Inter C N R S Ecol Geol de l'Himalaya* **268**, 181–191 (1977)
13 Frank, W, Gansser, A & Trommsdorff, V *Schweiz miner petrogr Mitt* **57**, 89–113 (1977)
14 Stoliczka, F *Mem geol Surv Ind* **5**, 12–15 (1865)
15 Norin, E *Rep Sino-Swedish Expedn Publ* **293**, 340 (1946)
16 Bhandari *et al Abstr 9th Sem Himalayan Geology* (1978)
17 Sharma, K K & Kumar, S *Him Geol* **8**, 262–287 (1978)
18 Sharma, K K & Gupta, K R *Rec Res Geol* **7**, 133–143 (1978)
19 Thakur, V C *et al J geol Soc Ind* **22**, 46–50 (1981)
20 Desio, A, Tongiorgi, E & Ferrara, G *22nd int Geol Cong* **11**, 479–497 (1964)
21 Molnar, P & Tapponnier, P *Science* **189**, 419 (1975)
22 Sharma, K K *et al Him Geol* **8**, 268–295 (1978)
23 Rai, H *Abstr 10th Sem Himalayan Geology* (1978)

Tightly bound β-hydroxy acids in a Recent sediment

Kimitaka Kawamura* & Ryoshi Ishiwatari

Department of Chemistry, Faculty of Science, Tokyo Metropolitan University, Fukuzawa 2-1-1, Setagaya-ku, Tokyo 158, Japan

The identification of β-hydroxy acids in Recent sediments[1–6] is considered to be evidence of a bacterial contribution to sedimentary organic matter. In previous studies, β-hydroxy acids have been extracted by saponification with KOH/methanol under reflux. We have now found that β-hydroxy acids are also present in Recent sediment in a tightly bound form which cannot be extracted by ordinary saponification. Release of those β-hydroxy acids requires that the sediment be heated to ~150 °C before saponification, or the use of harsher saponification.

* Present address Institute of Geophysics and Planetary Physics, University of California at Los Angeles, Los Angeles, California 90024, USA

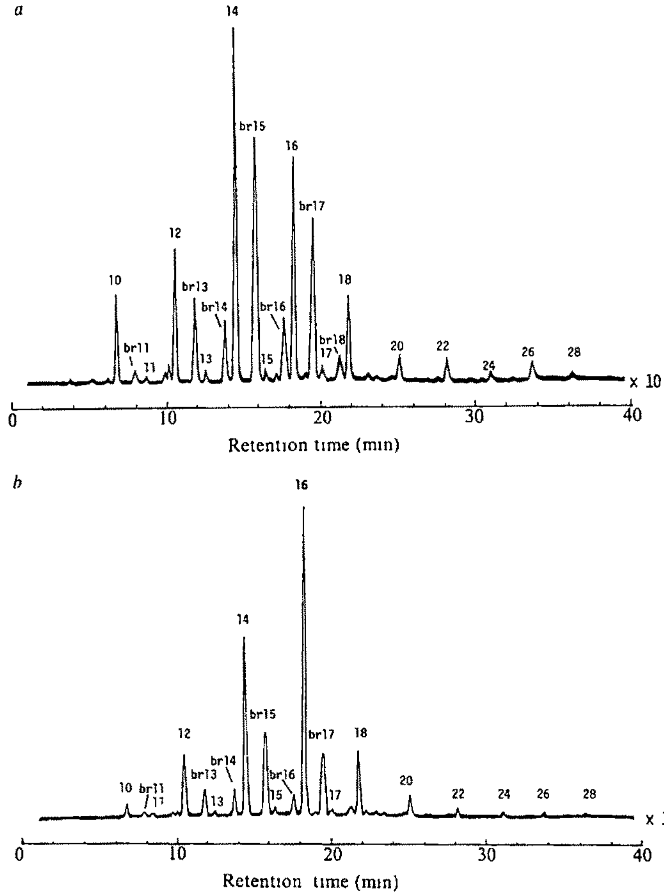


Fig. 1 Mass fragmentogram (m/z 175) of β-hydroxy acid methyl ester TMS ethers from unheated (a) and heated (154 °C, b) sediment samples

The use of mild heating to study the organic matter content of Recent sediments has been shown to increase the amount of extractable fatty acids[7–9] The increase is due to so-called 'tightly bound' fatty acids[9], which are thought to be located in organic and/or inorganic matrices where saponification reagents cannot easily act, or to be tightly linked in some unknown way with the organic and/or inorganic matrices[9] These studies suggested that organic molecules other than fatty acids might also be present in sediments in a 'tightly bound form' As β-hydroxy acids are considered markers of bacterial activity[1–4], it would be of interest to identify a tightly bound form of these compounds in sediment

Surface sediment samples were taken from Lake Biwa, the largest lake (altitude 85 m, area 674 4 km^2, maximum depth 102 m) in Japan, which is oligotrophic A wet sediment sample was homogenized using a Nihon-seiki (AM-1) homogenizer and aliquots of the sample (5–10 g wet wt) were heated at various temperatures (68–325 °C) in a sealed Pyrex tube (12 mm i d × 15 cm) under nitrogen for 24 h The unheated and heated samples were saponified with 50 ml of 0 5 M KOH/methanol containing 5% water under reflux for 2 h Saponification was carried out only once because we found that repeating the procedure increases the amount of β-hydroxy acids obtained by <3% The alkaline solution was extracted with 20 ml of n-hexane/ether (9 1) to remove neutral components The remaining solution was then acidified with concentrated HCl and the acid components extracted with n-hexane/ether (9 1) The extracts were methylated with 14% BF_3/methanol and the methyl esters separated into monocarboxylic acid fraction with n-hexane/benzene (9 1) and hydroxy acid fraction with benzene/ethyl acetate (1 1) by silica gel chromatography (Mallinckrodt 100 mesh, 5 mm i d × 10 cm) The latter fraction was treated with TMS-BA [N,O-Bis(trimethylsilyl)acetamide 25% in acetonitrile Tokyo Kasei Co Ltd] reagent to derive TMS ethers of hydroxy acid methyl

0028-0836/82/190144—02$01 00

Table 1 Concentrations of β-hydroxy acids released from the sediment, pre-extracted sediment and isolated kerogen

	Sediment*	Pre-extracted sediment	Kerogen
µg per g dry weight sediment	36.3	28.7	6.8

* The amount was obtained by subtracting the concentration (17.2 µg per g) for unheated sample from the concentration (53.5 µg per g) for heated sample at 154 °C.

esters, which were then analysed by gas chromatography–mass spectrometry (Shimadzu-LKB 9000) on a glass column (3 mm i.d.×2 m) packed with 1% OV-1, with programmed temperature increases of 5 °C per min from 100 to 280 °C. The β-hydroxy acids were quantified by comparing the peak heights in the mass fragmentograms scanned at m/z 175, which is a characteristic fragment of β-trimethylsilyloxy methyl esters[1,2], with authentic standards (β-hydroxy normal C_{14}, C_{16} and C_{18} acids).

β-Hydroxy C_{10}–C_{28} acids were detected in the unheated and heated samples. Mass fragmentograms scanned at m/z 175 for the unheated and heated samples (Fig. 1) showed that most even carbon-numbered β-hydroxy acids are unbranched (normal), whereas odd carbon-numbered acids are almost entirely iso/ante-iso branched chain compounds. A similar distribution has been reported[3] for bound β-hydroxy acids from algal mats sediments. As predicted, the amount of both normal and branched β-hydroxy acids increased on heating; in particular, the concentration of normal β-hydroxy C_{16} acid increased by as much as fivefold. With increasing temperature, the total amount of β-hydroxy acids increases threefold and maximizes at 150–180 °C (Fig. 2). As sedimentary β-hydroxy acids are considered to be produced by bacteria[1–4], our observations suggest the presence in the sediment of bacteria which produce specifically a large amount of β-hydroxy n-C_{16} acid. Apart from n-C_{16} acid, the relative proportions of constituents in the unheated and heated samples are similar, suggesting that other acids have a similar origin.

β-Hydroxy acids could not have been formed directly through β-oxidation of saturated fatty acids because we confirmed that authentic n-C_{17} saturated fatty acid did not generate any β-hydroxy acids when heated with wet Ca-montmorillonite at 160 °C for 24 h, and because we could not detect any decrease in the concentration of saturated fatty acids in sediment heated to 150–180 °C. α,β-Unsaturated fatty acids are other possible precursors of β-hydroxy acids. We consider this route unlikely, however, because α,β-unsaturated fatty acids have never been found in sediments[10,11]. It therefore seems that the β-hydroxy acids released on heating may not be artefacts but may have existed originally in the sediments.

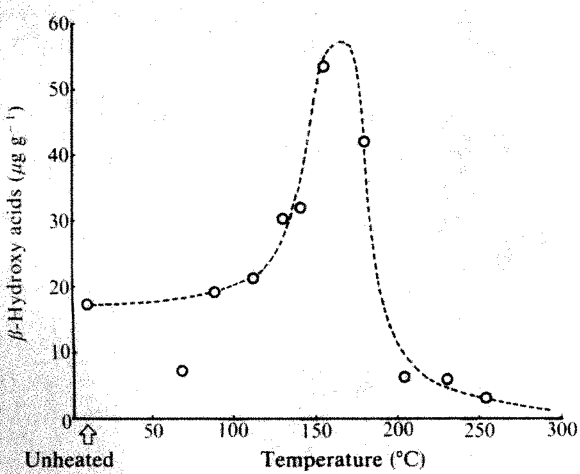


Fig. 2 Changes in the concentrations of β-hydroxy acids in sediments on heating.

Next, we conducted a harsher saponification (180 °C, 3 h) of the pre-extracted sediment (0.5 g) and the isolated kerogen (34.5 mg, ash content 32.5%) using a Pyrex tube (12 mm i.d.× 15 cm) containing 5 ml of 2 M KOH solution[12]. As shown in Table 1, we obtained a considerable amount of β-hydroxy acids. The amount of β-hydroxy acids released from the pre-extracted sediment by harsher saponification (28.7 µg g^{-1}) is almost equal to that released from the sediment by heating (36.2 µg g^{-1}). β-Hydroxy acids were also released from kerogen, although at a lower concentration (6.8 µg g^{-1} of dry sediment) than from the pre-extracted sediment. Harsher saponification with 2 M KOH solution may dissolve and/or decompose some humic materials, amorphous silica and clay minerals. It is therefore likely that most β-hydroxy acids were associated with those materials: that is, 'tightly bound' β-hydroxy acids can be considered to be occluded by (and partly linked with) humic materials, including kerogen and amorphous silica, and to be trapped in the interlayers of clay minerals. Further work is necessary to define the type of binding.

β-Hydroxy acids have been found in bacteria[2,5,13–16], but not in algae. We did not detect these compounds in phytoplankton (*Closterium*) taken from Lake Biwa. Therefore, it is obvious that sedimentary β-hydroxy acids are a result of bacterial activity. As most (68%) of the β-hydroxy acids in the sediment exist in a 'tightly bound form', the quantities of β-hydroxy acids reported previously may be underestimates.

This study was supported in part by a grant from the Ministry of Education, Japan.

Received 25 August 1981; accepted 8 March 1982.

1. Eglinton, G., Hunneman, D. H. & Douraghi-Zadeh, K. *Tetrahedron* **24,** 5929–5941 (1968).
2. Boon, J. J., DeLange, F., Schuyl, P. J. W., DeLeeuw, J. W. & Schenk, P. A. in *Advances in Organic Geochemistry 1975* (eds Campos, R. & Goni, J.) 255–272 (Enadimsa, Madrid, 1977).
3. Cardoso, J. N., Eglinton, G. & Holloway, P. J. in *Advances in Organic Geochemistry 1975* (eds Campos, R. & Goni, J.) 273–287 (Enadimsa, Madrid, 1977).
4. Perry, G. J., Volkman, J. K., Johns, R. B. & Bavor, H. J. Jr *Geochim. cosmochim. Acta* **43,** 1715–1725 (1979).
5. Volkman, J. K., Johns, R. B., Gillan, F. T., Perry, G. J. & Babor, H. J. Jr *Geochim. cosmochim. Acta* **44,** 1133–1143 (1980).
6. Cranwell, P. A. *Geochim. cosmochim. Acta* **45,** 546–552 (1981).
7. Baedecker, M. J., Ikan, R., Ishiwatari, R. & Kaplan, I. R. in *Chemistry of Marine Sediments* (ed. Yen, T. F.) 55–72 (Ann Arbor Science, Michigan, 1977).
8. Harrison, W. E. *Chem. Geol.* **21,** 314–334 (1978).
9. Kawamura, K. & Ishiwatari, R. *Geochem. J.* **15,** 1–8 (1981).
10. Van Vleet, E. S. & Quinn, J. G. *Nature* **262,** 126–128 (1976).
11. Matsuda, H. & Koyama, T. *Geochim. cosmochim. Acta* **41,** 341–345 (1977).
12. Schnitzer, M. & Neyrould, A. *Fuel* **54,** 17–19 (1975).
13. Moss, C. W., Samuels, S. B., Liddle, J. & McKinney, R. M. *J. Bact.* **114,** 1018–1024 (1973).
14. Mayberry, W. R., Smith, P. F., Langworthy, T. A. & Plackett, P. *J. Bact.* **116,** 1091–1095 (1973).
15. Boon, J. J., DeLeeuw, J. W., Hoek, G. J. & Vosjan, J. H. *J. Bact.* **129,** 1183–1191 (1977).
16. Mayberry, W. R. *J. Bact.* **143,** 582–587 (1980).

Surface coatings on ancient coccoliths

Peter M. Bürki*, Lesley S. Dent Glasser & Douglas N. Smith

Department of Chemistry, University of Aberdeen, Meston Walk, Old Aberdeen AB9 2UE, UK

Coccoliths are rings of calcite plates, often of intricate design, produced by unicellular algae[1] (Fig. 1). Still widespread in modern oceans, these algae must have been extremely abundant in the Cretaceous period, and their skeletal remains form the main component of natural chalk. Both modern and ancient coccoliths are well preserved even in the corrosive environments of seawater and sediments, and the surface properties of natural chalk are said to be different from those of synthetic calcite. It has been suggested that this may be the result of a surface membrane; the present study establishes the existence of a coating on ancient coccoliths, and discusses its nature.

* Present address: Holderbank Management & Consulting Ltd, Materials Division, CH-5113 Holderbank, Switzerland.

0028-0836/82/190145—03$01.00

It has been demonstrated[2,3] that modern coccoliths are surrounded by a thin film of organic material. This could be composed of polysaccharides and amino acids[4] and may be formed when the coccolith is produced within the algal cell[5]. By careful dissolution of Oligocene coccoliths on electron microscope grids, Hamano and Honjo[6] observed residual membranes which they believed to be composed partly of organic matter. They found that the acid-insoluble residue from the coccolith fraction of the chalk contained polysaccharides, amino acids and clay-like material. Taylor (unpublished work) suggested that thin layers of precipitated clay minerals might cover and protect coccoliths in chalk. This postulate is substantiated by the work of Weir and Catt[7] and Jeans[8] who concluded that smectite is precipitated from porewater in chalk sediments. Suess[9] has also postulated organo-clay associations for all pre-Recent limestones.

In the present study, samples of Cretaceous chalk from Swanscombe were gently decalcified on electron microscope

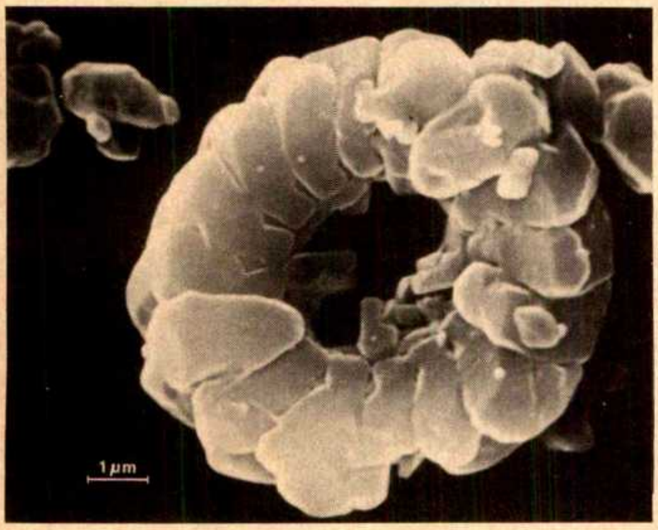


Fig. 1 Scanning electron micrograph of a Cretaceous coccolith. This is only one of the many forms observed.

grids and examined in a CORA analytical transmission electron microscope (Kratos Ltd, AEI Scientific Instruments). Residues of decalcified coccoliths are shown in Figs 2 and 3. These complete 'skins' are clearly replicas of the original structures, an example of which is shown in Fig. 1. Slower dissolution in EDTA buffered with NH_4Cl/NH_4OH produced both partly and completely decalcified rings, confirming that the coatings are on the surface of the coccoliths. Whole skins, however, were rather rare; distorted fragments either partly retaining the original structure or lacking any recognizable form were more common. The coatings may collapse after dissolution or be ruptured by the production of CO_2.

Electron microanalysis was difficult because the coatings are very thin (a few hundred ångstroms at most) giving a low count rate. By examining areas over holes in the substrate, the problem of background count was avoided, and the approximate atomic ratios found were: 1.0 Si : 0.1 Mg : 0.3 Al : 0.04 K : 0.1 Fe. P, from apatite crystals, S and Cl were also occasionally encountered. Significant levels of Ca were present only when visible fragments of calcite remained. These ratios, which were the same with all three dissolution procedures (see legend to Fig. 2), are typical of a smectite, but none of the skins gave a diffraction pattern. Similar ratios were found for the other amorphous debris of ill-defined form. Crystalline clay minerals, quartz and apatite were also observed in the residues, occasionally attached to the coatings (Fig. 3).

Because the microanalyser cannot detect carbon, no conclusion can be drawn about the organic nature of these skins.

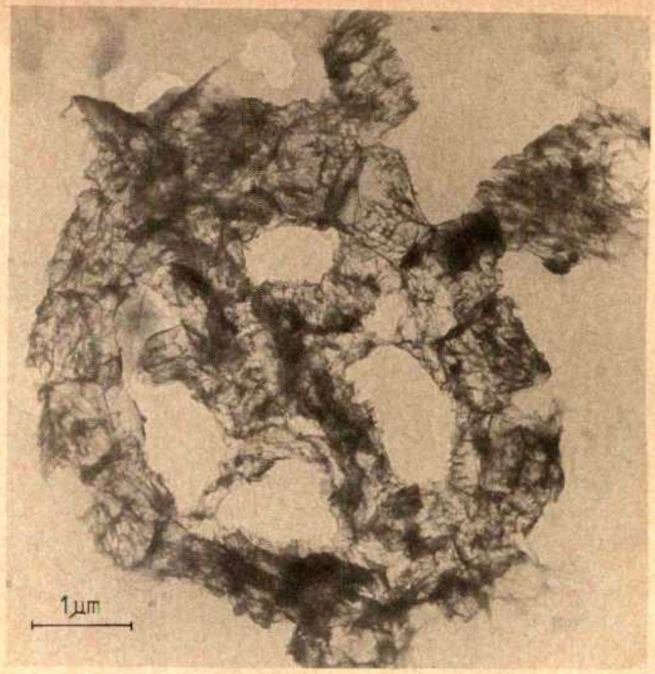


Fig. 2 Decalcified Cretaceous coccolith. The calcite was removed by floating the specimen, mounted on an electron microscope grid, on a 0.005 M EDTA solution for 30 s, then washing in distilled water for 2 h. Slower and more controllable dissolution was achieved by incorporating 5% NH_4Cl/NH_4OH buffer (pH 10) in the EDTA. Removal of the calcite with a dilute sodium acetate/acetic acid buffer (pH 5) is also possible but tends to rupture the skins. In the example shown the coating on the central 'bridge' feature, often found in coccoliths, has remained intact.

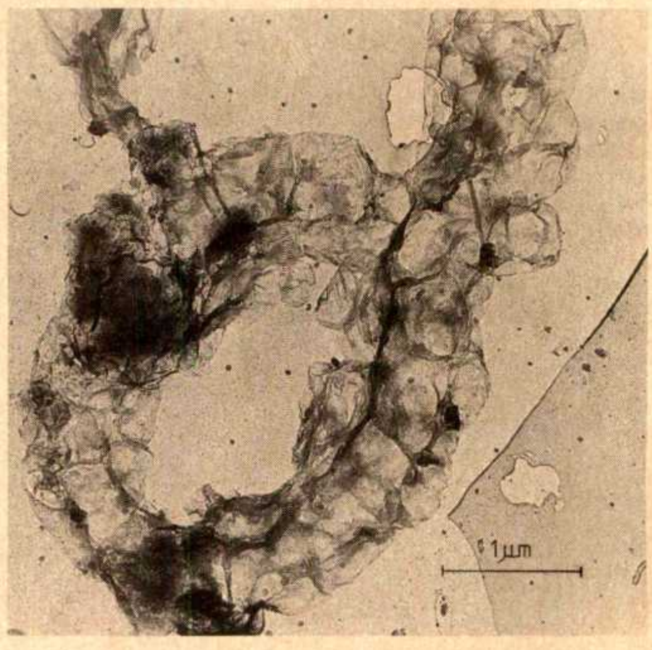


Fig. 3 Decalcified Cretaceous coccolith. The dark area on the left is smectite. The rest of the skin is amorphous.

However, C and H analysis of the chalk before and after treatment with bromine water shows that 68% of the total organic carbon is unavailable for oxidation. This could be because it is incorporated in the coatings as suggested by Hamano and Honjo[6], possibly bedded between the calcite and the amorphous alumino-silicate layer.

From the present evidence we cannot say whether the inorganic part of the coatings was incorporated during their formation or formed subsequently by preferential precipitation of siliceous material dissolved in the mobile porewater[8]. However this may be, the accurate replicas shown in Figs 2 and 3 could plainly be silicified organic membranes, and demonstrate unequivocally that Cretaceous coccoliths can have a coating closely associated with the calcite.

We thank the Welwyn Hall Research Association for financial support for this work.

Received 25 January; accepted 18 March 1982.

1. Hay, W. W. & Mohler, H. P. *J. Paleont.* **41**, 1505–1541 (1967).
2. Pautard, F. G. E. *Biological Calcification; Cellular and Molecular Aspects* (ed. Schraer, H.) 140–141 (North-Holland, Amsterdam, 1970).
3. Klaveness, D. thesis, Oslo Univ. (1971).
4. de Jong, L. W., Dam, W., Westbroek, P. & Crenshaw, M. A. in *The Mechanisms of Mineralisation in the Invertebrates and Plants* (eds Watabe, N. & Wilbur, K. M.) 135–153 (University of South Carolina Press, 1976).
5. Hay, W. W. in *Geology and Archaeology of Northern Cyrenaica, Libya* (ed. Barr, F. T.) 149–157 (Petroleum Exploration Society of Libya, 1968).
6. Hamano, M. & Honjo, S. *J. geol. Soc. Japan* **75**, 607–614 (1969).
7. Weir, A. H. & Catt, J. A. *Clay Miner.* **6**, 97–110 (1965).
8. Jeans, C. V. *Clay Miner.* **7**, 311–329 (1968).
9. Suess, E. *Geochim. cosmochim. Acta* **34**, 157–159 (1970).

Flow-field variables trigger landing in flies

Hermann Wagner
Max-Planck-Institut für biologische Kybernetik, Spemannstrasse 38, D-7400 Tübingen 1, FRG

The eyes of moving animals experience a continuously changing flow of stimulation. Information contained in this 'optical flow field' can be extracted to control movement[1–4]. I present here evidence that pre-landing deceleration in flies is determined by such information: onset of deceleration is triggered when the ratio of the image expansion of a target on the retina to the image size ('relative retinal expansion velocity', RREV) reaches a critical value.

Female houseflies (*Musca domestica* L.) flying freely in a cage ($90 \times 70 \times 100$ cm) were simultaneously filmed from two sides at 50, 80 or 100 frames per s while landing on black stationary spheres (radius, $r = 0.25$–1.0 cm). Frame-by-frame analysis allowed the reconstruction of three-dimensional landing trajectories. A sample stereogram is shown in Fig. 1. I evaluated certain variables, including (1) the fly's velocity (v), defined as the distance travelled along the flight path between two frames, divided by the sampling time Δt; (2) the distance from the fly to the centre of the target (x); (3) the change of distance with time ($\Delta x/\Delta t$, which is the component of the velocity in the direction of the target); (4) the solid angle subtended by the target on the retina (retinal size a); and (5) the target's retinal expansion velocity ($\Delta a/\Delta t$). As the retinal image is the only source of optical information directly available to the animal, the physical variables (1–3) can only be derived indirectly via the optical variables (4, 5). x, r and a are related as follows[5]:

$$a = 2\pi(1-\sqrt{1-(r/x)^2}) \approx \pi r^2/x^2 \quad (\text{for } x \gg r) \qquad (1)$$

(if $x = 2r$, the error of the approximation is 7%) and in analogy to Lee[15] for the approximation:

$$\text{RREV} = \frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}t} = -2\frac{1}{x}\frac{\mathrm{d}x}{\mathrm{d}t} \quad \text{or} \quad \text{RREV} = \frac{\mathrm{d}}{\mathrm{d}t}\ln a = -2\frac{\mathrm{d}}{\mathrm{d}t}\ln x \qquad (2)$$

Flies reduce their flight velocity before landing and the onset of the final deceleration before landing is taken to define the beginning of the landing phase. Figure 2*a* shows the velocity profile of different landing trajectories compared by setting zero time to the onset of deceleration. Tests were carried out to determine whether the onset is related to the variables x, a, $\Delta a/\Delta t$, $1/a \times \Delta a/\Delta t$ and v/x. Table 1 shows data obtained from 31 landing trajectories, pooled at corresponding times to allow calculation of the coefficients of variation (c.v.) for each variable. As the c.v. is a measure of dispersion defined by standard deviation/mean, variables showing a high c.v. cannot be regarded as critical for the initiation of pre-landing deceleration. At each sampling point the c.v. of both the RREV and the velocity/distance quotient are much smaller than all others, which suggests the prime importance of these variables in initiating onset of deceleration. As velocity and distance are only indirectly accessible, it is concluded that the critical cue is the RREV and not v/x.

The RREV is given by the ratio of retinal expansion velocity to retinal size (equation 2). By correlating these variables it

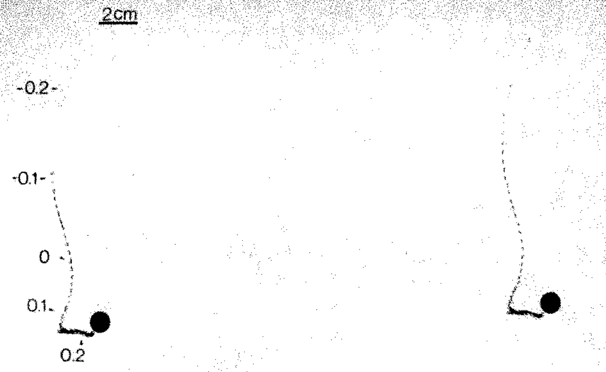


Fig. 1 Example of a fly trajectory viewed from above. It should be viewed with normal stereoglasses. Open circles and lines represent the head and long axis of the fly, respectively. The solid circle indicates the position and size of the target. The legs could not be seen on the films. Numbers represent the time (s) with onset of deceleration set to zero time; 100 frames per s.

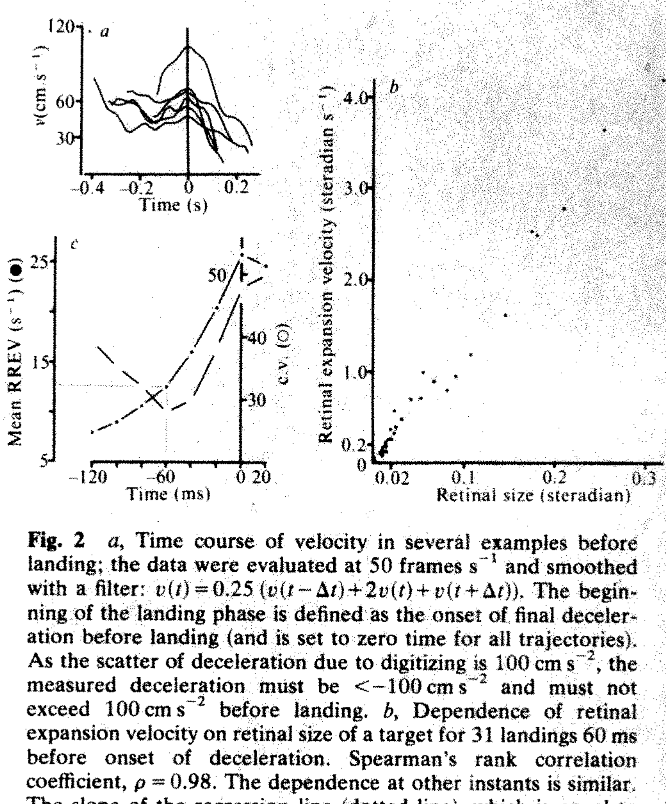


Fig. 2 *a*, Time course of velocity in several examples before landing; the data were evaluated at 50 frames s^{-1} and smoothed with a filter: $v(t) = 0.25\,(v(t-\Delta t)+2v(t)+v(t+\Delta t))$. The beginning of the landing phase is defined as the onset of final deceleration before landing (and is set to zero time for all trajectories). As the scatter of deceleration due to digitizing is 100 cm s^{-2}, the measured deceleration must be < -100 cm s^{-2} and must not exceed 100 cm s^{-2} before landing. *b*, Dependence of retinal expansion velocity on retinal size of a target for 31 landings 60 ms before onset of deceleration. Spearman's rank correlation coefficient, $\rho = 0.98$. The dependence at other instants is similar. The slope of the regression line (dotted line), which is equal to the RREV, is 13.2 s^{-1}; the high correlation coefficient demonstrates that the variance of the RREV is low. *c*, Mean (●) and c.v. (○) of the RREV at different time shifts with respect to onset of deceleration (zero time). Coefficient of variation data for the RREV are the same as in Table 1. The c.v. minimum at $t = -60$ ms may indicate the reaction time from the threshold RREV to the onset of deceleration. Dotted lines show the value of the RREV corresponding to −60 ms.

0028-0836/82/190147—02\$01.00

Table 1 Coefficients of variation for different variables with respect to onset of final deceleration

Time (ms)	+20	0	−20	−40	−60	−80	−100	−120
Distance (x)	78	72	66	62	58	55	53	52
Retinal size (a)	128	124	128	127	128	128	137	143
Retinal expansion velocity ($\Delta a/\Delta t$)	143	132	133	128	128	119	113	109
RREV ($1/a \times \Delta a/\Delta t$)	50	48	38	30	28	32	35	39
Velocity/distance (v/x)	48	36	29	27	28	30	32	32

At each instant only one value per trajectory was used to calculate the c.v.. The c.v. of the RREV and the quotient of velocity and distance from fly to target should be identical if the flies approach the target in a straight line [equation (2)]. The differences between these c.v. indicate that this condition is often not fulfilled.

was shown that during the last 120 ms before onset of the final deceleration their dependence on each other was strong, that is, the variance of the RREV was small, confirming the conclusion drawn from Table 1. This is shown in Fig. 2*b* at 60 ms before onset of deceleration, the instant of maximum dependence corresponding to the minimum of the c.v. of the RREV (Fig. 2*c*, open circles). The mean value of the RREV increases steadily (Fig. 2*c*, solid circles). Together with the correlation analysis, this finding leads to the conclusion that, for onset of deceleration to occur, the RREV must exceed a critical value. The minimum of the c.v. at −60 ms (Fig. 2*c*, open circles) might represent the response delay (see ref. 6). Thus, the threshold value of the RREV is $\sim 13\ \mathrm{s}^{-1}$ (dotted lines in Fig. 2*c*).

Previous experiments on landing using tethered flying insects[7–13] did not yield the same conclusions as presented here from free-flight experiments. In the earlier experiments, extension of the legs before landing ('landing response'[7]), which could not be seen on my films, was taken as a criterion for the initiation of the landing procedure. The conclusions from these experiments are consistent neither amongst themselves nor—partly—with the findings presented here. In particular, a critical retinal target size alone, as proposed by Eckert and Hamdorf[10], does not initiate landing of free-flying flies as shown by the high c.v. (Table 1). Furthermore, as the RREV is almost independent of real target size (equations 1 and 2), the present conclusion can better explain landings on targets of various dimensions than hypotheses which take into account retinal size or retinal expansion velocity alone. Other authors[8,11–13] found that the landing response could be elicited by centrifugal movement of striped patterns. This simulates target expansion and these results are thus qualitatively consistent with those reported here. However, the dependence of the landing response on the RREV has not yet been investigated, either in experiments with tethered flying flies or in electrophysiological investigations of units with selective response to an expanding stimulus[14].

Formally, the RREV is the inverse of 'time to collision', that is, the time needed to reach a target directly approached at constant velocity[15]. As many landing trajectories are neither direct nor of constant velocity, a simple interpretation of the RREV as time to collision is not applicable. The RREV, however, represents the more general case: it gives information about the distance to a target not in absolute, but in relative units dependent on both the translation velocity and the direction of flight with respect to a target. This means that a distance is tagged with a time. Thus, the important cues for landing are times and not distances. Similar arguments might hold for escape responses and even control of cruising flights. The present conclusions agree well with a theory established by Lee and co-workers[4,15] for humans and birds, and prove that exploiting variables in the optical flow field can provide simple solutions for rather complex orientation tasks.

I thank Professor W. Reichardt for discussions and helpful criticism. Drs J. Brede, H. Bülthoff, E. Buchner, T. Collett, M. Egelhaaf, K. Hausen, T. Poggio, K. Vogt, Chr. Wehrhahn and J. Zeil critically read an early draft of the paper and made many useful suggestions. I thank Dr R. Cook for correcting the English. Computer programs were modified from a program package developed by H. Bülthoff, Chr. Koch and T. Poggio.

Received 21 December 1981; accepted 26 March 1982.

1. Wehner, R. in *Handbook of Sensory Physiology* Vol. 7/6C (ed. Autrum, H.) 287–616 (Springer, Berlin, 1981).
2. Reichardt, W. & Poggio, T. *Q. Rev. Biophys.* **9**, 311–346 (1976).
3. Collett, T. S. *J. comp. Physiol.* **138**, 271–282 (1980).
4. Lee, D. N. & Reddish, P. E. *Nature* **293**, 293–294 (1981).
5. Wagner, H. *Diplomarbeit der Universität Tübingen* (1980).
6. Land, M. F. & Collett, T. S. *J. comp. Physiol.* **89**, 331–357 (1974).
7. Goodman, L. J. *J. exp. Biol.* **37**, 854–878 (1960).
8. Coggshall, J. C. *J. exp. Biol.* **57**, 401–413 (1972).
9. Eckert, H. *Biol. Cybern.* **37**, 235–247 (1980).
10. Eckert, H. & Hamdorf, K. *J. comp. Physiol.* **138**, 253–264 (1980).
11. Perez de Talens, A. F. & Taddei-Ferretti, C. *J. exp. Biol.* **52**, 233–256 (1970).
12. Taddei-Ferretti, C. & Perez de Talens, A. F. *Z. Naturf.* **28c**, 568–592 (1973).
13. Wehrhahn, C., Hausen, K. & Zanker, J. *Biol. Cybern.* **41**, 91–99 (1981).
14. Chillemi, S. & Taddei-Ferretti, C. *J. exp. Biol.* **94**, 105–118 (1981).
15. Lee, D. N. *Perception* **5**, 437–459 (1976).

A peripheral locus for amphetamine anorexia

Michael G. Tordoff, Jill Hopfenbeck, Larry L. Butcher & Donald Novin

Department of Psychology and Brain Research Institute, University of California, Los Angeles, California 90024, USA

The anorexic action of amphetamine is thought to be mediated by central dopaminergic and noradrenergic neurotransmitter systems[1–4]. However, some of its sympathomimetic effects can be attenuated by denervation[5] or by depletion of peripheral noradrenaline[6,7]. This led us to consider the possibility that a component of amphetamine anorexia might be due to excitation of the sympathetic innervation of the viscera. To test this, the amphetamine dose–response curves of free-feeding rats with visceral sympathetic denervation by coeliac ganglionectomy (GAN) were compared with those of sham-operated controls (SHM). We report here that ganglionectomized rats were protected from the anorexic effects of low doses of amphetamine (0.2 or 0.4 mg per kg interaperitoneal, i.p.). We suggest that the anorexia might result from enhanced hepatic glycogenolysis[8], and the subsequent stimulation of hepatic metabolic receptors that inhibit feeding[9].

Sixteen male Long-Evans hooded rats (Charles River) weighing 216–246 g were maintained on a 12 : 12 h light/dark cycle (dark period beginning at 17.00) with food (Wayne lab blox, Chicago) and water continuously available. Half the animals were subjected to coeliac ganglionectomy by the method of Lambert[10]; the others were subjected to identical procedures except that the ganglion was not excised. After surgery, GAN rats showed a transient 2-day hypophagia accompanied by diarrhoea, which is a good indication of complete ganglionectomy. Rats were given 24 days to recover after surgery. They were then tested in groups of four (two GAN and two SHM) every four days. All received four saline injection trials, to habituate them to the drug administration and measurement regimen, followed by counterbalanced injections of saline or (+)amphetamine sulphate (0.2, 0.4, 0.6, 0.8 or 1.0 mg per kg i.p.).

The procedure used for habituation and testing was as follows: 1 h before the start of the dark period, each rat was

0028–0836/82/190148—03$01.00

weighed and placed in a cage designed to measure activity by photobeam crossings[11]. Food was available on the cage floor and tap water could be obtained from a drinking tube. Just before the lights were turned off, each rat was injected and a fresh quantity of food was provided. Food intake was measured after 30, 60, 90 and 120 min by weighing the remaining food and subtracting this value from its initial weight. Activity was monitored continuously but, for purposes of data analysis, was grouped into 30-min intervals.

Separate three-way analyses of variance (group × dose × time), followed by Tukey's *post hoc* tests, were done on the food intake and activity measurements. Amphetamine caused a dose-related anorexia ($F(5,70)=8.31$, $P<0.001$): during the first 90 min after injection, the food intake of SHM rats was significantly depressed by all doses of amphetamine. However, GAN rats showed no decrease in consumption in response to amphetamine at either 0.2 or 0.4 mg per kg ($F(5,70)=2.42$, $P<0.05$; see Fig. 1*a*). During the same interval, both groups showed a comparable dose-related hyperactivity ($F(5,70)=28.59$, $P<0.001$; see Fig. 1*b*). Thus, the feeding dose–response curves of SHM and GAN rats differed, but their activity responses were the same. Disrupting the sympathetic innervation of the viscera caused an attenuation of anorexia that was independent of the hyperkinetic action of amphetamine. This sympathetic mechanism accounted for a relatively large proportion of the anorexia at low doses of amphetamine, and relatively little at higher doses.

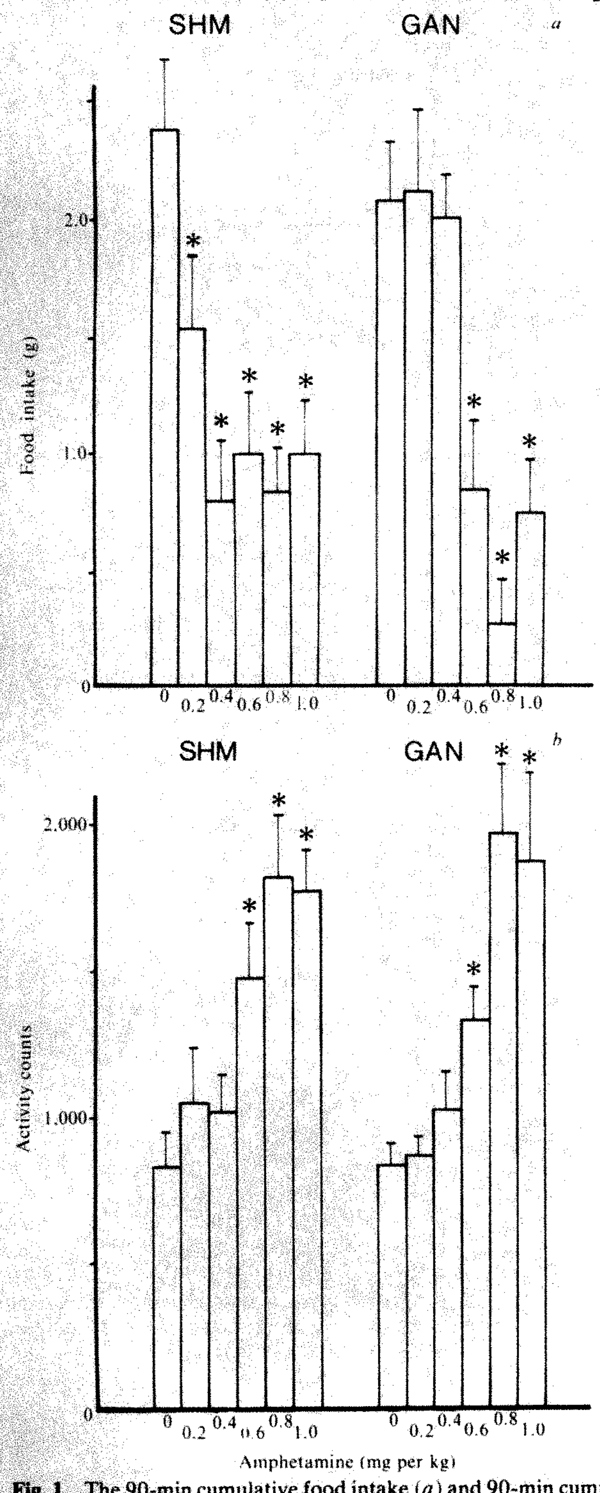


Fig. 1 The 90-min cumulative food intake (*a*) and 90-min cumulative activity (*b*) of sham (SHM) and coeliac ganglionectomized (GAN) rats were measured after administration of various doses of amphetamine. * $P<0.05$ compared with saline-injected controls (Tukey's test).

It is unlikely that nonspecific effects of the coeliac ganglionectomy affected the results. The food intake of SHM and GAN rats after injections of saline was statistically indistinguishable, as was activity after all injections. Furthermore, GAN rats ate more food than controls after low doses of amphetamine, which suggests there was no general debilitation.

The coeliac ganglion innervates most of the abdominal viscera. However, we consider the most likely site for the peripheral anorexic action of amphetamine to be the liver. The liver has considerable sympathetic innervation[12] and is involved in the glucostatic regulation of feeding. Russek[9] postulated that hepatic receptors sensitive to glucose or one of its metabolites respond to the liberation of glucose from glycogen, so that promotion of glycogenolysis stimulates hepatic receptors, which inhibit feeding. In support of this hypothesis, glucose, 2-deoxy-D-glucose and adrenaline have greater effects, with shorter latency, when administered intraportally or intraperitoneally than systemically[8,13,14]. In addition, vagotomy attenuates the anorexia induced by glucose, glucagon, adrenaline and amphetamine[14–17]. Both electrical stimulation of hepatic sympathetic fibres and i.p. injection of amphetamine induce glycogenolysis[8,18]. Finally, glucagon and amphetamine are more potent in rats with high rather than low glycogen levels[8,17]. A possible mechanism for peripherally-induced amphetamine anorexia could therefore involve the stimulation of sympathetic hepatic nerve fibres, hence inducing glycogenolysis and subsequent activation of metabolic receptors capable of inhibiting feeding[19,20].

Because the coeliac ganglion has both efferent and afferent hepatic innervation, including a large component of the vagus nerve, this study did not determine whether amphetamine acts either directly on the liver or its local innervation, or indirectly via a sympathetic discharge which originates in the CNS and which activates sympathetic fibres that travel to the liver. However, recent evidence suggests that direct stimulation of the liver by amphetamine may be relatively unimportant, as amphetamine-induced anorexia was more effectively blocked by coeliac ganglionectomy than by subdiaphragmatic vagotomy but adrenaline-induced anorexia was attenuated to the same extent by these two surgical procedures[16]. As adrenaline is a sympathomimetic which, unlike amphetamine, does not easily enter the brain, and because the coeliac ganglion provides hepatic sympathetic innervation, the additional protection from amphetamine anorexia afforded by coeliac ganglionectomy may reflect interruption of sympathetic discharge of CNS origin.

Blundell and Burridge[21] suggested that amphetamine anorexia consists of additive noradrenergic and dopaminergic components. In their model, the anorexia attributable to noradrenaline (NA) increases over low doses of amphetamine but reaches a maximum ~0.5 mg per kg. The anorexia attributable to dopamine (DA) has little effect in this range, but increases rapidly for doses >0.5 mg per kg. Although there is evidence which seems to contradict the Blundell and Burridge model[22–24], it is tempting to speculate that the dual anatomical (brain and periphery) and neurochemical (DA and NA) mechanisms may correspond. One might represent a centrally-mediated dopaminergic factor, and the other, a peripherally-mediated noradrenergic sympathetic discharge factor. Our

finding that GAN and SHM animals showed differences only with doses of amphetamine <0.5 mg per kg provides support for such a speculation, as the Blundell and Burridge model predicts this to be the dose where the major contribution to amphetamine anorexia changes from NA to DA.

Previous reports have negated the possibility of a peripheral locus for the anorexia induced by amphetamine, but these studies have used large doses, food-deprived animals or insensitive measuring techniques[2,3]. We have found that these conditions are the reverse of those required to observe differences in feeding between sympathetically denervated subjects and their controls. This suggests that the usual techniques for screening anorexic drugs using large doses and food-deprived rats may be inappropriate, because such tests would not recognize active anorexigens that had sympathetico-hepatic effects.

This research was supported by National Institute of Health NS07687 to D. Novin and NS10928 to L.L.B.

Received 20 November 1981; accepted 18 March 1982.

1. Cole, S. O. *Psychol. Bull.* **79**, 13–20 (1973).
2. Harris, S. C., Ivy, A. C. & Searle, L. M. *J. Am. med. Ass.* **134**, 1468–1475 (1947).
3. Nathanson, M. H. *J. Am. med. Ass.* **108**, 528–531 (1937).
4. Leibowitz, S. F. *Brain Res.* **98**, 529–545 (1975).
5. Lockett, M. F. *Br. J. Pharmac.* **5**, 485–496 (1950).
6. Burn, J. H. & Rand, M. J. *J. Physiol., Lond.* **144**, 314–336 (1958).
7. Muscholl, E. A. *Rev. Pharmac.* **6**, 107–128 (1966).
8. Russek, M. & Stevenson, J. A. F. *Physiol. Behav.* **8**, 245–249 (1972).
9. Russek, M. *Nature* **197**, 79–80 (1963).
10. Lambert, R. *Surgery of the Digestive System in the Rat*, 484 (Charles Thomas, Springfield, Illinois, 1965).
11. Hodge, G. K. & Butcher, L. L. *Pharmac. Biochem. Behav.* **10**, 695–705 (1979).
12. Williams, P. L. & Warwick, R. *Functional Neuroanatomy of Man*, 1068–1083 (Churchill-Livingstone, Edinburgh, 1975).
13. Russek, M., Lora-Vilchis, M. C. & Islas-Chaires, M. *Physiol. Behav.* **24**, 157–161 (1972).
14. Novin, D., Vanderweele, D. A. & Rezek, M. *Science* **181**, 858–860 (1973).
15. Novin, D., Sanderson, J. & Gonzalez, M. *Physiol. Behav.* **22**, 107–113 (1979).
16. Tordoff, M. G., Novin, D. & Russek, M. *J. comp. physiol. Psychol.* **96**, 361–375 (1982).
17. Vanderweele, D. A., Geiselman, P. J. & Novin, D. *Physiol. Behav.* **23**, 155–158 (1979).
18. Shimazu, T. & Fukuda, A. *Science* **150**, 1607–1608 (1965).
19. Mogenson, G. J. *Physiol. Behav.* **3**, 133–136 (1968).
20. Russek, M. & Bruni, E. *Ann. Inst. Univ. Auton. Mexico, 41, Ser. Biol. Exp.* **1**, 17–24 (1970).
21. Blundell, J. E. & Burridge, S. L. in *The Treatment of Obesity* (ed. Munro, J. F.) 53–83 (University Park Press, Baltimore, 1979).
22. Franklin, K. B. J. & Herberg, L. J. *Neuropharmacology* **16**, 45–46 (1977).
23. Frey, H.-H. & Schulz, R. *Biochem. Pharmac.* **22**, 3041–3049 (1973).
24. Schmitt, H. *J. Pharmac. Paris* **4**, 285–294 (1973).

Effect of chloroform on charge movement in the nerve membrane

J. M. Fernández*, F. Bezanilla* & R. E. Taylor*†

* Department of Physiology, Ahmanson Laboratory of Neurobiology and Jerry Lewis Neuromuscular Research Center, University of California, Los Angeles, California 90024, USA and The Marine Biological Laboratory, Woods Hole, Massachusetts 02543, USA
† Laboratory of Biophysics, National Institute of Neurological and Communicative Disorders and Stroke, National Institutes of Health, Bethesda, Maryland 20205, USA and The Marine Biological Laboratory, Woods Hole, Massachusetts 02543, USA

It is generally accepted that membrane-bound charges or dipoles moving under the influence of the transmembrane electric field produce conformational changes that result in the opening and closing of ion-selective channels[1]. The location of these charged groups or dipoles is unknown but it has been speculated that they are sited either inside the channel macromolecule or embedded in the lipid bilayer[2]. We report here that in the squid axon membrane, the general anaesthetic chloroform reduces the magnitude of the gating currents but has no effect on their kinetics. In contrast, chloroform increases the rate of charge translocation of the lipophilic ion dipicrylamine (DpA⁻). These results indicate that the gating charges responsible for the opening and closing of the Na^+ channel do not move in the bulk lipid of the membrane as do DpA^- molecules, but that they move in a physical domain where their kinetics are unaffected by changes in the structural parameters of the lipid matrix. It is therefore unlikely that general anaesthetics act to modify the kinetics of the opening and closing of ionic channels via modifications of the structural parameters of the lipid matrix[3].

Experiments were performed on internally perfused and voltage-clamped squid giant axons using the methods of Bezanilla *et al.*[4] to record Na^+ and gating currents. Figure 1*a* shows that perfusion of the axon with a solution containing 60 mM chloroform reduced dramatically the magnitude of the inward Na^+ current. To observe sodium and gating currents simultaneously, the axon was bathed in a reduced Na^+ concentration (Fig. 1*b*, trace 1). In these conditions, 43 mM of chloroform added to the internal perfusion solution reduced the Na^+ current in parallel with the gating current (Fig. 1*b*, trace 2). Both effects were reversed on removing the chloroform (Fig. 1*b*, trace 3). The effect of chloroform on the Na^+ and gating currents was not reversed by hyperpolarizing prepulses, suggesting that these effects are not related to the mechanisms of inactivation.

To analyse the effects of chloroform on gating currents, the Na^+ current was eliminated by substituting the external Na^+ with the impermeant ion Tris and by the addition of 300 nM tetrodotoxin (TTX). Figure 1*b*, traces 4, 5, shows that

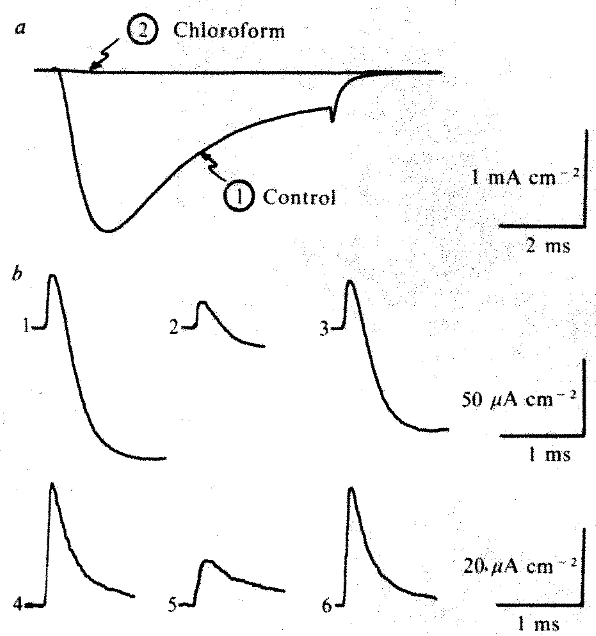


Fig. 1 Effect of chloroform on the Na^+ inward current and the gating currents. *a*, Trace 1, inward Na^+ current before treatment with chloroform; trace 2, abolition of the Na^+ current when the axon was internally perfused with a solution containing 60 mM chloroform. Traces 1 and 2 were obtained under voltage clamp by pulsing to −10 mV from a holding potential (HP) of −70 mV. Leakage and linear capacitive currents were subtracted using the P/4 technique with a subtracting holding potential (SHP) of −140 mV (for details of the technique see ref. 4). The external solution was 440 mM NaCl, 50 mM $MgCl_2$, 10 mM $CaCl_2$, *p*H 7.2. The internal solution was 100 mM *N*-methylglucamine(NMG) glutamate, 100 mM NMG fluoride and ~400 mM sucrose adjusted to make the internal and external solution isosmotic, *p*H 7.2, 5 °C. *b*, Trace 1, axon with low external Na^+ allowed measurement of the ON gating current as well as the Na^+ inward current. Trace 2, when perfused with a solution containing 43 mM chloroform, a significant reduction in the Na^+ current was paralleled by a reduction in the gating current. Trace 3, on removal of chloroform, both effects were reversed. Trace 4, ON gating currents measured in the same axon as in traces 1–3, after eliminating the Na^+ current with an external solution containing Tris and 300 nM TTX. Trace 5, partial decrease in the size of the ON gating current when the axon was perfused with the same solution as in trace 2. Trace 6, after removing chloroform, the gating current was almost completely recovered. The full effect of chloroform on the Na^+ and gating currents occurred within seconds of beginning the perfusion with chloroform-containing solutions; recovery was complete within <1 min after removal of chloroform. Traces 1–6 were obtained by pulsing to 30 mV from HP = −70 mV using the P/4 procedure with a SHP of −140 mV. The external solution (traces 1–3) was 88 mM NaCl, 360 mM Trizma 7.2, 50 mM $MgCl_2$, 10 mM $CaCl_2$. Traces 4–6 were obtained using an external solution containing 450 mM Trizma 7.2, 50 mM $MgCl_2$, 10 mM $CaCl_2$ and 300 nM TTX. The internal solution in each case was 100 mM NMG fluoride, 100 mM NMG glutamate and ~400 mM sucrose, *p*H 7.2, 5 °C. Where indicated, chloroform was added to the internal solution, which was stirred for 30 min in a closed glass cylinder. The chloroform concentrations correspond to the fraction (v/v) added. In the experiments reported here we did not add chloroform to the external solution because the experimental chamber consisted of acrylic plastic (Perspex).

0028-0836/82/190150—03$01.00

chloroform reduced the size of the gating current. As expected, this effect was reversible (Fig. 1*b*, trace 6).

Gating currents during a depolarizing or hyperpolarizing pulse (ON gating) can be described by the sum of two voltage-dependent exponential relaxations[5] (Fig. 2*a*, trace 1); the solid line indicates a two-exponential fit with time constants $\tau_F = 126\ \mu s$ and $\tau_S = 571\ \mu s$. When the axon was internally perfused with a solution containing 60 mM chloroform, the remaining gating current (Fig. 2*a*, trace 2) could again be fitted with two exponentials having time constants $\tau_F = 112\ \mu s$ and $\tau_S = 587\ \mu s$ which, within our resolution, are identical to those obtained in the fit of trace 1. However, the amplitude of the fast exponential component was reduced by about half but there was no significant change in the amplitude of the slow exponential component. This effect was seen for both depolarizing and hyperpolarizing pulses. The difference record between the gating current before addition of chloroform (Fig. 2*a*, trace 1) and during chloroform (trace 2), shown in Fig. 2*b*, is fitted by a single exponential (solid trace) with a time constant $\tau_F = 126\ \mu s$. Figure 2*b* represents the fraction of the fast component eliminated by chloroform. A more dramatic reduction of the fast exponential component is shown in Fig. 2*c*.

The main effect of chloroform was to decrease the amplitude of the fast component of gating with no significant effect on the kinetics of either the remaining fast or slow components. These results were qualitatively reproduced in several experiments, with variations only in their magnitude (see Fig. 2*a*, *c*) due to our present inability to control accurately the concentration of chloroform. The magnitude of the effect obtained was dependent on internal and external perfusion rates, therefore the actual concentration of chloroform reaching the membrane was probably significantly smaller than the values given throughout this paper. The concentrations given were estimated from the fraction of chloroform (v/v) added to the internal perfusion solution.

Several models have been proposed for the action of general anaesthetics such as chloroform on the structure of lipid bilayers. These include modifications of such structural parameters as thickness, dielectric constant, dipole potential and the degree of order (for example, fluidity) of the lipid hydrocarbon chains (see refs 3, 6). Lipophilic ions such as DpA^- and tetraphenylboron have proved useful in elucidating these structural parameters of lipid bilayers[3,7]. The general behaviour of lipophilic ions in membranes is phenomenologically similar to that of gating currents, because they bear a charge but are trapped inside the membrane[7]. It has been shown that changes in the structure and composition of the bilayer greatly affect the transport of these lipophilic ions. For example, Reyes and Latorre[3] demonstrated that pharmacological concentrations of the neutral general anaesthetics chloroform and benzyl alcohol significantly increase the translocation rate of the hydrophobic ion tetraphenylborate in lipid bilayers, indicating that these anaesthetics change the dielectric constant of the lipid matrix. As these physical changes should affect the movement of any membrane-bound charge, a similar effect would be expected for the charge movement of the gating particles responsible for the opening and closing of Na^+ channels. However, the results presented here show that this is not the case. The gating current kinetics of the fast and slow components were unmodified by chloroform, which indicates that the mechanism of action of chloroform on gating currents is very different from that on the charge movement induced by lipophilic ions.

To confirm that chloroform has the same general effect on the membrane of squid giant axon as it does on lipid bilayers, we incorporated the lipophilic ion DpA^- into the squid giant axon membrane and measured the effect of chloroform on the induced charge movement. As shown in Fig. 3*a*, DpA^- produced a very large charge movement having characteristics similar to those described by Benz and Conti[8] (details of these results will be published elsewhere). When the axon was

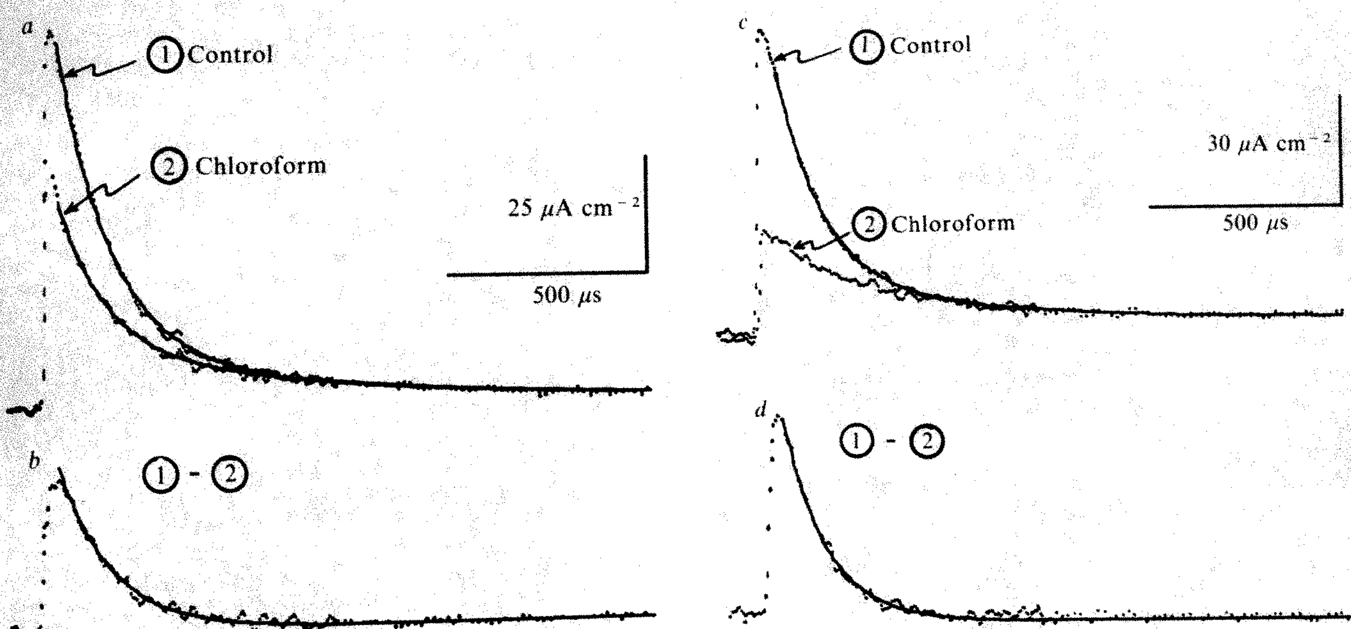


Fig. 2 Effect of chloroform of the ON gating currents. *a*, Trace 1 shows an ON gating current measured using the P/4 procedure in the absence of ionic currents for a test pulse of +100 mV from a holding potential of −70 mV (SHP = −150 mV). The solid line represents a two-exponential fit with a fast time constant, $\tau_F = 126\ \mu s$ and a slow time constant, $\tau_S = 571\ \mu s$. Trace 2 shows the effect of internally perfusing the axon with a solution containing 60 mM chloroform; this trace could again be fitted with two exponentials having time constants $\tau_F = 112\ \mu s$ and $\tau_S = 587\ \mu s$. Chloroform reduced the amplitude of the fast component to about half without altering the kinetics of the remaining fast and slow components. *b*, Numerical difference between traces 1 and 2 of *a*, which corresponds to the fraction of the fast component eliminated by chloroform. The solid line represents a single exponential fit with time constant $\tau_F = 126\ \mu s$. *a*, A more dramatic reduction of the fast component for a different axon internally perfused with a solution containing 60 mM chloroform. Trace 1 corresponds to the ON gating current for a test pulse of +110 mV from a holding potential of −70 mV (SHP = −140 mV). A two-exponential fit (solid line) gave the time constants $\tau_F = 108\ \mu s$ and $\tau_S = 508\ \mu s$. Trace 2 shows the gating current remaining when the axon was perfused with the chloroform-containing solution; in this case the fast component of trace 1 was reduced by ~85%. *d*, The numerical difference between traces 1 and 2 in *c*; the solid line corresponds to a single exponential fit with a time constant $\tau_F = 111\ \mu s$.

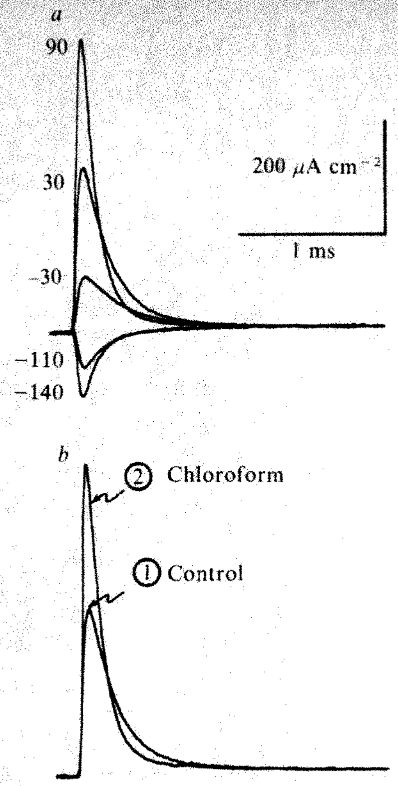


Fig. 3 Charge movement induced by the lipophilic ion dipicrylamine in the absence of ionic currents measured using the P/4 procedure. *a*, Charge movement measured after perfusing the axon internally and externally with solutions containing 10^{-5} M DpA^- for ~5 min. The currents correspond to pulses applied, from a holding potential of −70 mV, to the indicated membrane potential. As the induced charge movement was so large, correction for the normal gating charge movement was unnecessary. *b*, Effect of internally perfusing the axon with a solution containing 43 mM chloroform on the kinetics of the DpA^--induced charge movement. In contrast to the effect observed for the ON gating current, the rate of DpA^--induced charge movement was increased by chloroform (traces 1 and 2 were obtained by pulsing to +30 mV from a HP of −70 mV with a SHP of −140 mV). The effect was reversed after removal of chloroform. The data shown are for the same axon and solutions as used for the axon of Fig. 1*b* (traces 4–6), and demonstrate that the contrasting effects of chloroform on the Na^+ gating currents and the DpA^--induced charge movement can be observed in the same axon in identical conditions.

perfused with a solution containing 43 mM chloroform, the translocation rate of DpA^- was increased significantly as indicated by a larger initial current and a faster decay (Fig. 3*b*). The voltage dependence of the DpA^- charge movement was unmodified by chloroform, indicating that the observed change in the translocation rate represented a reduction in the activation energy that the DpA^- molecules must overcome to cross the membrane, in good agreement with the observations of Reyes and Latorre[3] for artificial bilayers. Thus chloroform has similar effects on the structural parameters of squid axon and lipid bilayer membranes.

These experiments therefore indicate that Na^+ channel gating charges do not diffuse through the bulk lipid matrix of the membrane as the DpA^- molecules do. It is therefore questionable to use hydrophobic ions as analogues of the Na^+ gating charge movement. This conclusion has important implications as it renders unlikely those classes of model in which the gating charges of the Na^+ channel move freely in the lipid matrix and in which gating is due to the aggregation of several charged proteins that are oriented by the field to form an operational channel[2]. These experiments also rule out models in which the action of general anaesthetics is to enhance the rate at which the Na^+ gating charges move between closed and open states, subsequently altering excitability[3]. The results suggest that chloroform decreases the number of operational channels by preventing the movement of the charge that produces the fast component of gating current.

The major difficulty in the study of gating currents has been the inability to examine the various components separately. Our experiments suggest the possibility of using chloroform as a tool to dissociate the fast from the slow component and to study the correlation between these components and the Na^+ conductance to elucidate further the gating processes of voltage-dependent channels.

We thank Drs S. Krasne, J. Vergara and M. White for helpful comments on the manuscript. This work was supported by a Grass Fellowship to J.M.F. and by USPHS grant GM 30376.

Received 4 December 1981; accepted 16 March 1982.

1. Armstrong, C. M. & Bezanilla, F. *Nature* **242**, 459–461 (1973).
2. Mueller, P. *Ann. N.Y. Acad. Sci.* **264**, 247–263 (1976).
3. Reyes, J. & Latorre, R. *Biophys. J.* **28**, 259–280 (1979).
4. Bezanilla, F., Taylor, R. E. & Fernández, J. M. *J. gen. Physiol.* **79**, 21–40 (1982).
5. Bezanilla, F. & Taylor, R. E. *Biophys. J.* **23**, 479–484 (1979).
6. Franks, N. P. & Lieb, W. R. *J. molec. Biol.* **133**, 469–500 (1979).
7. Szabo, G. *Ann. N.Y. Acad. Sci.* **303**, 266–280 (1977).
8. Benz, R. & Conti, F. *J. Membrane Biol.* **59**, 91–104 (1981).

Regulation of muscarinic ligand binding sites by nerve growth factor in PC12 phaeochromocytoma cells

J. E. Jumblatt* & A. S. Tischler†

* Department of Biochemistry and Pharmacology and
† Department of Pathology, Tufts University School of Medicine, 136 Harrison Avenue, Boston, Massachusetts 02111, USA

The rat phaeochromocytoma cell line PC12 (ref. 1) possesses many characteristics of adrenal chromaffin cells, including the ability to synthesize, store and release catecholamines in response to acetylcholine and other depolarizing agents[2–4]. When exposed to 2.5S nerve growth factor (NGF), PC12 cells cease to divide, and exhibit a number of neurone-like characteristics, including the extension of neurites, development of voltage-sensitive ion channels and electrical excitability, and increased capacity to synthesize and release acetylcholine[3–8]. Biochemical and pharmacological evidence indicates that both adrenal chromaffin cells and sympathetic neurones contain, in addition to ganglionic-type nicotinic receptors, muscarinic cholinergic receptors proposed to modulate catecholamine release[9–13]. Catecholamine release from PC12 cells is directly stimulated by nicotinic, but not muscarinic, cholinergic agonists[3]. Although ion flux through acetylcholine-activated channels in PC12 cells is inhibited by high concentrations of the potent muscarinic antagonist quinuclidinyl benzilate (QNB)[14], the possibility that these cells may contain functional muscarinic receptors has been largely overlooked. We show here that PC12 cells contain saturable, high-affinity binding sites for ^{3}H-(−)-quinuclidinyl benzilate (^{3}H-QNB), which seem pharmacologically comparable with muscarinic receptors described in rodent brain[15,16], smooth muscle[17,18] and adrenal medulla[12]. Moreover, treatment with NGF produces a marked elevation of specific ^{3}H-QNB binding. PC12 cells thus may provide a useful system for investigating the biochemical mechanisms, cellular distribution and regulation of muscarinic receptors by hormones and trophic signals.

PC12 cells (passages 23–32) were derived from the clone originally isolated by Greene and Tischler[1]. Cells were cultured in collagen-coated 10-cm polystyrene dishes as described previously[1,6]. To treat cells with NGF for binding experiments, replicate cultures were inoculated at a density of 4×10^4 cells per cm^2 (except where otherwise indicated), and cultured for

0028–0836/82/190152—03$01.00

12–16 days in control medium or in medium supplemented with 50 ng ml^{-1} 2.5S NGF (ref. 19). The medium was changed every third day. To collect cells, the culture medium was replaced by 5 ml Ca^{2+}, Mg^{2+}-free phosphate-buffered saline[20] (CMF-PBS) containing 1 mM EDTA, and cells were gently removed using a rubber policeman. Complete detachment of cell bodies and neurites was monitored by phase microscopy. Cells were washed twice by centrifugation (1,000*g*, 10 min) and resuspension in 0.1 vol. of 320 mM sucrose, 50 mM phosphate, 0.1 mM EDTA *p*H 7.2 (SPE buffer) at 4 °C, and counted using a haemacytometer. To prepare membranes, cells were homogenized for 30 s with a Potter-Elvehjem homogenizer and centrifuged for 10 min at 1,000*g* to pellet the nuclei. Homogenization and centrifugation were repeated on the nuclear pellet, and the combined supernatants were centrifuged at 120,000*g* for 60 min to obtain a crude membrane (mitochondria + microsomes) fraction. The membrane pellet was resuspended in SPE buffer (3–5 mg protein per ml) and used within 12 h for the binding measurements. All fractionation steps were performed at 0–4 °C.

The concentration dependence of ^{3}H-QNB binding at equilibrium to PC12 cell membranes was determined in the absence and presence of 10^{-6} M atropine (Fig. 1). Scatchard analysis[21] of specific ^{3}H-QNB binding (Fig. 1*b*; total minus nonspecific binding) indicated a single, high-affinity class of binding sites having a $K_d = 0.146 \pm 0.07$ nM, $B_{max} = 32.7 \pm 7.2$ fmol per mg protein. These results are in close agreement with the reported affinity of QNB binding to muscarinic cholinergic receptors in rat brain[15], smooth muscle[17,18] and adrenal medulla[11,12]. Interestingly, the density of QNB binding sites in PC12 cell membranes is comparable with that reported for rat adrenal medulla membranes[12], but more than 10-fold higher than in bovine adrenal medulla[11]. Assuming 100% recovery of receptor-containing membranes, PC12 cells have ~1,400 specific QNB binding sites per cell. By comparison, Patrick and Stallcup[22] reported these cells to contain 5,000 binding sites per cell for

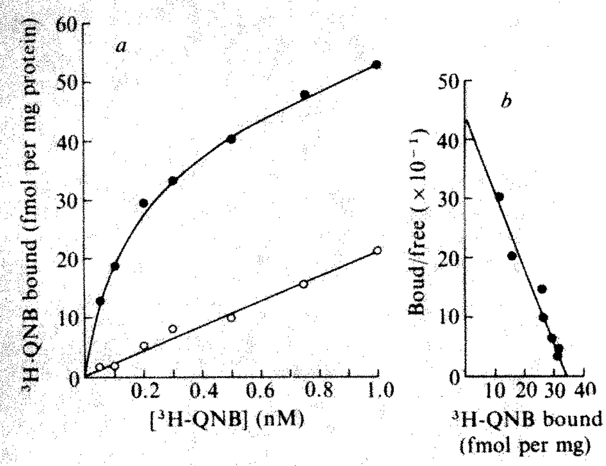


Fig. 1 Binding of ^{3}H-QNB to membranes from untreated PC12 cells as a function of ^{3}H-QNB concentration. PC12 cell membranes (100–160 μg protein) were preincubated for 10 min at 30 °C in microfuge tubes containing 0.9 ml of 50 mM $Na^+ + K^+$ phosphate buffer *p*H 7.2 plus or minus atropine sulphate (10^{-6} M final concentration). A volume of ^{3}H-QNB (NEN; 31.2 Ci mmol^{-1}) was then added to each tube to give the indicated final ^{3}H-QNB concentrations. Following incubation at 30 °C for 60 min, 0.8-ml aliquots were removed from each sample and filtered under suction through 25 mM Whatman GF/F glass fibre filters. Each filter was washed four times with 5 ml of ice-cold 50 mM phosphate, then dried and counted in 10 ml Beckman Ready Solv HP scintillation fluid using a Packard Tri-Carb liquid scintillation counter. All points are the means of duplicate determinations, which deviated by <15% from the indicated values. The protein content was determined by the Lowry method[24]. *a*, Total binding of ^{3}H-QNB (●); nonspecific binding in the presence of 10^{-6} M atropine (○). *b*, Scatchard plot of specific ^{3}H-QNB binding from the experimental data shown in *a*.

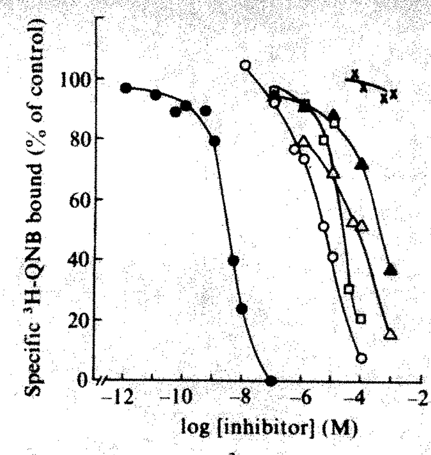


Fig. 2 Competition for specific ^{3}H-QNB binding to PC12 membranes by different concentrations of cholinergic drugs: atropine (●), oxotremorine (○), pilocarpine (□), (+)tubocurarine (▲), carbamylcholine (△), nicotine (×). Binding assays were performed essentially as described in Fig. 1 legend except that different concentrations of competing ligand were present during the initial preincubation and a single concentration of ^{3}H-QNB (0.2–0.6 nM) was present during the final incubation. Nonspecific binding has been subtracted from total binding and per cent of maximum specific binding in controls without inhibitor has been plotted. All points represent the means of duplicate or triplicate measurements, which ranged within 15% of the means.

^{125}I-α-bungarotoxin—a selective probe of nicotinic cholinergic receptors at neuromuscular junctions. The relationship of α-bungarotoxin binding sites to functional nicotinic receptors in PC12 cells is unclear, however, because the toxin fails to block the nicotinic receptor-mediated responses in these cells[22].

To evaluate the pharmacological specificity of ^{3}H-QNB binding, various cholinergic agents were tested at several concentrations for inhibition of specific QNB binding at a fixed concentration of radioligand (Fig. 2). Corrected IC_{50} values obtained from these data are listed in Table 1. The order of potency (atropine > oxotremorine > pilocarpine > carbamylcholine > (+)tubocurarine > nicotine) and the IC_{50} values for these agents correspond closely to the pharmacological selectivity of ligand binding to muscarinic cholinergic receptors in brain and other tissues[11,15–18].

Muscarinic receptors exhibit complex binding characteristics for agonists (Hill coefficients < 1), whereas binding of antagonists usually conforms to a simple mass action isotherm (Hill coefficients ~1)[23]. The basis for this complexity is unknown, but may reflect receptor heterogeneity or conformational shifts[23]. Hill analysis (not shown) of the inhibition data in Fig. 2 suggested similar complexity for agonist but not antagonist binding to PC12 membranes. Estimated Hill coefficients were 0.55 for carbamylcholine, 0.74 for oxotremorine, 0.81 for pilocarpine and 1.06 for atropine.

NGF treatment of PC12 cells is known to induce choline acetyltransferase, increased synthesis, storage and release of

Table 1 Relative abilities of cholinergic drugs to inhibit specific ^{3}H-QNB binding to PC12 cell membranes

Drug	IC_{50}*(M)
Atropine	1.2×10^{-9}
Oxotremorine	1.6×10^{-6}
Pilocarpine	7.4×10^{-6}
Carbamylcholine	2.6×10^{-5}
(+)Tubocurarine	1.5×10^{-4}
Nicotine	$> 10^{-3}$

Values are derived from the data presented in Fig. 2. IC_{50}* is the concentration of inhibitory drug which reduced specific ^{3}H-QNB binding by 50%, corrected for the shift in radioligand occupancy using the formula $IC_{50}^* = IC_{50}/1 + [L]/K_d$ where K_d is the measured equilibrium dissociation constant for ^{3}H-QNB (0.146 nM) and $[L]$ is the concentration of radioligand present (0.2–0.6 nM).

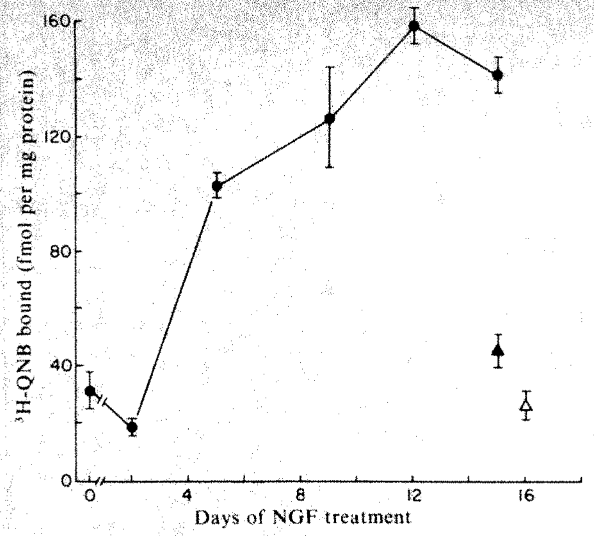


Fig. 3 Effect of NGF treatment on specific ^{3}H-QNB binding. Parallel cultures were plated and maintained in medium with or without NGF (50 ng ml^{-1}). Medium was changed every third day. Specific binding was determined as described in Fig. 1 legend, using 10^{-9} M ^{3}H-QNB in the presence and absence of 10^{-6} M atropine. Values represent means ± s.e.m. of triplicate determinations. NGF-treated cells plated at a density of 4×10^4 cells per cm^2 (●); untreated cultures plated at a density of 4×10^4 cells per cm^2 (▲); untreated cultures plated at 2×10^3 cells per cm^2 (△).

acetylcholine, elevation of acetylcholinesterase activity and the appearance of clustered vesicular structures resembling cholinergic synaptic vesicles[1,6-8]. When we investigated the effects of NGF, on QNB binding, we found that the level of specific ^{3}H-QNB binding sites in PC12 membranes increased markedly during 15 days of NGF treatment, following an apparent lag of 2–5 days (see Fig. 3). Mean values for specific QNB binding were 156.5 ± 7.1 fmol per mg protein after 12–15 days of NGF treatment compared with 28.3 ± 6.9 fmol per mg for untreated cells. The increase depended on the presence of NGF and was absent in untreated cultures maintained for the same period at high cell densities (Fig. 3). Interestingly, the major increase in binding, evident by day 5 (Fig. 3), preceded extensive neurite outgrowth.

This study using PC12 cells provides the first evidence for upward regulation of muscarinic ligand binding sites by a chemically defined, developmentally important, neurotrophic substance. These binding sites appear pharmacologically equivalent to muscarinic receptors in normal neural tissues. Their physiological functions both in PC12 cells and in normal tissues, however, require further study. Muscarinic receptors are proposed to influence adrenal cells and sympathetic neurones in several ways[9-13], but the most extensive physiological data are derived from the mammalian heart *in vivo*, where presynaptic muscarinic receptors on sympathetic nerve endings augment the effects of vagus stimulation by inhibiting release of catechol- amines[9]. PC12 cells thus may provide a useful system to study the biochemical aspects of muscarinic receptor function *in vitro*, as well as the effects of hormones and trophic substances on the expression of these receptors.

We thank Val Slayton for technical assistance. This work was supported by grants from the American Cancer Society (PDT-171) and the National Cancer Institute, DHHS (R23 CA-27808), and the Charles A. King Trust.

Received 18 August 1981; accepted 12 March 1982.

1. Greene, L. A. & Tischler, A. S. *Proc. natn. Acad. Sci. U.S.A.* **73**, 2424–2428 (1976).
2. Greene, L. A. & Rein, G. *Brain Res.* **129**, 247–263 (1977).
3. Greene, L. A. & Rein, G. *Brain Res.* **138**, 521–528 (1977).
4. Dichter, M. A., Tischler, A. S. & Greene, L. A. *Nature* **268**, 501–504 (1977).
5. Ritchie, A. K. *J. Physiol., Lond.* **286**, 541–561 (1979).
6. Tischler, A. S. & Greene, L. A. *Lab. Invest.* **39**, 77–89 (1978).
7. Greene, L. A. & Rein, G. *Nature* **268**, 349–351 (1977).
8. Schubert, D., Heinemann, S. & Kidokoro, Y. *Proc. natn. Acad. Sci. U.S.A.* **74**, 2579–2583 (1977).
9. Löffelholz, K. & Muscholl, E. *Naunyn-Schmiedebergs Arch. exp. Path. Pharmak.* **265**, 1–15 (1969).
10. Sharma, V. K. & Banerjee, S. P. *Nature* **272**, 276–278 (1978).
11. Kayaalp, S. O. & Neff, N. H. *Neuropharmacology* **18**, 909–911 (1979).
12. Kayaalp, S. O. & Neff, N. H. *Eur. J. Pharmac.* **57**, 255–257 (1979).
13. Adnan, N. A. M. & Hawthorne, J. N. *J. Neurochem.* **36**, 1858–1860 (1981).
14. Stallcup, W. B. *J. Physiol., Lond.* **286**, 525–540 (1979).
15. Yamamura, H. I. & Snyder, S. H. *Proc. natn. Acad. Sci. U.S.A.* **71**, 1725–1729 (1974).
16. Kloog, Y., Egozi, Y. & Sokolovsky, M. *Molec. Pharmac.* **15**, 545–558 (1979).
17. Yamamura, H. I., Wastick, G., Charg, K. J. & Snyder, S. H. *Proc. West. Pharmac. Soc.* **19**, 13–18 (1976).
18. Yamamura, H. I. & Snyder, S. H. *Molec. Pharmac.* **10**, 861–867 (1974).
19. Bocchini, V. & Angelletti, P. U. *Proc. natn. Acad. Sci. U.S.A.* **64**, 787–794 (1969).
20. Dulbecco, R. & Vogt, M. *J. exp. Med.* **99**, 167–182 (1954).
21. Scatchard, G. *Ann. N.Y. Acad. Sci.* **51**, 660–672 (1949).
22. Patrick, J. & Stallcup, B. *J. biol. Chem.* **252**, 8629–8633 (1977).
23. Birdsall, N. J. M. & Hulme, E. C. *J. Neurochem.* **27**, 7–16 (1976).
24. Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. *J. biol. Chem.* **193**, 265–275 (1951).

Adenosine diphosphate induces binding of von Willebrand factor to human platelets

Takayoshi Fujimoto & Jack Hawiger*

Departments of Pathology and Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, USA

Blood platelets seal off leaks in the vessel wall to prevent haemorrhage by a mechanism involving their activation, adhesion to the zone of vessel injury and aggregation[1]. This physiologically important haemostatic process requires Factor VIII/von Willebrand factor (vWF), a plasma glycoprotein missing in the genetic, haemorrhagic disorder called von Willebrand disease, in which platelets cannot adhere to the subendothelial layer of the vessel wall[2]. Thus vWF is necessary for anchoring platelets to the vessel wall, and it mediates platelet–platelet interaction. This interaction requires an appropriate receptor on the platelet surface. Until recently, the antibiotic ristocetin was used to induce *in vitro* receptor-mediated binding of vWF to platelets and to cause their aggregation or agglutination[3-7]. But although a useful laboratory tool[8], ristocetin is clearly non-physiological, so we turned our attention to thrombin, which appears in venous blood within 45 s of venipuncture and is a very potent physiological inducer of platelet aggregation. Thrombin causes platelets to change shape from discoid to spiny spheres, and alters the platelet membrane resulting in release of ADP, serotonin and other constituents of platelet storage granules[9]. We recently showed that low concentrations (0.1–0.5 nM) of thrombin, attainable in blood, induce binding of vWF to a specific platelet receptor[10]. We now report that the binding of vWF to platelet receptors is also induced by ADP on its addition to platelets. Moreover, ADP released from platelet storage granules by thrombin is, at least in part, responsible for subsequent binding of ^{125}I-labelled vWF.

Platelets from human blood were separated from plasma proteins by stepwise albumin gradient centrifugation and Sepharose 2B gel filtration as described earlier[11]. All experiments were performed with platelets suspended in HEPES buffer, pH 7.35, containing 0.1% dextrose and 0.35% albumin. Human vWF was purified and labelled with ^{125}I, as before[10]. This preparation was rendered essentially free of fibrinogen and fibronectin by passage through an anti-fibrinogen antibody Sepharose 4B affinity column and a gelatin–Sepharose 4B column. Purified vWF gave negative results for fibrinogen content in immunodiffusion and in the staphylococcal clumping test having a detection limit of 0.5 μg fibrinogen per ml[12]. By the use of an even more sensitive radioimmunoassay, we found that 1 mg of purified vWF contained < 60 ng of fibrinogen, that is, < 0.00006% (w/w).

Binding of ^{125}I-vWF to platelets occurred in 0.5 ml reaction mixture containing 1×10^8 platelets with 5 μM ADP or 0.5 nM

* To whom correspondence should be addressed.

0028–0836/82/190154—03$01.00

thrombin, to which hirudin and then ^{125}I-labelled vWF had been added. Incubation was at room temperature (22 °C) in polypropylene tubes (12×75 mm) without stirring. Incubation was terminated by rapid centrifugation (15,000 r.p.m., 2 min) of platelet suspension through 9:1 (v/v) mixture of dibutylphthalate and Apiezon Oil C (ref. 13) at 60 min, by which time binding had reached equilibrium (see below). The supernatant was sampled for free radioactivity, and bound radioactivity was measured in the platelet pellet which was cut off together with the tip of the tube after rapid freezing of its contents in a mixture of dry ice and acetone. Aggregation of platelets was measured photometrically in a Payton (Buffalo, New York) dual-channel aggregometer at 37 °C with constant stirring (900 r.p.m.) according to Born's method[14].

Intact resting platelets, separated from plasma proteins, bound little ^{125}I-vWF. After stimulation with 5 μM ADP, at least three times more ^{125}I-vWF was bound (Fig. 1); this binding approached saturation at 2.8–5.6 μg ^{125}I-vWF, and equilibrium binding was achieved within 30 min. Whereas approximately 75% of the bound ^{125}I-vWF could be displaced by a 100-fold excess of unlabelled vWF, a similar excess of unlabelled fibrinogen or fibronectin had no effect, suggesting that binding was specific for ^{125}I-vWF.

Induction of binding of ^{125}I-vWF to ADP-stimulated human platelets could be effectively abolished with either the ADP-splitting enzyme, apyrase (5.4 U; see Fig. 1), or an ADP-converting system comprising creatinine phosphate (7.5 mM) and creatinine phosphate kinase (3.0 U) (results not shown).

In the above experiments, exogenous ADP was added to platelets before ^{125}I-vWF. However, the dense granules of platelets are a rich source of ADP which is easily releaseable by thrombin[9]. We had earlier shown that thrombin (0.5 nM) induces binding of ^{125}I-vWF to human platelets[10] and so we decided to determine whether there was any relationship between thrombin-induced release of ADP and thrombin-induced binding. The finding that addition of the ADP-splitting enzyme apyrase to thrombin-treated platelets reduced binding by 73% showed that endogenous ADP, released from platelet dense granules and present in the platelet 'environment', was responsible in large part for inducing the binding of ^{125}I-vWF to thrombin-stimulated platelets. ADP-induced binding of ^{125}I-vWF required calcium as both EDTA and EGTA (2 mM) inhibited the interaction.

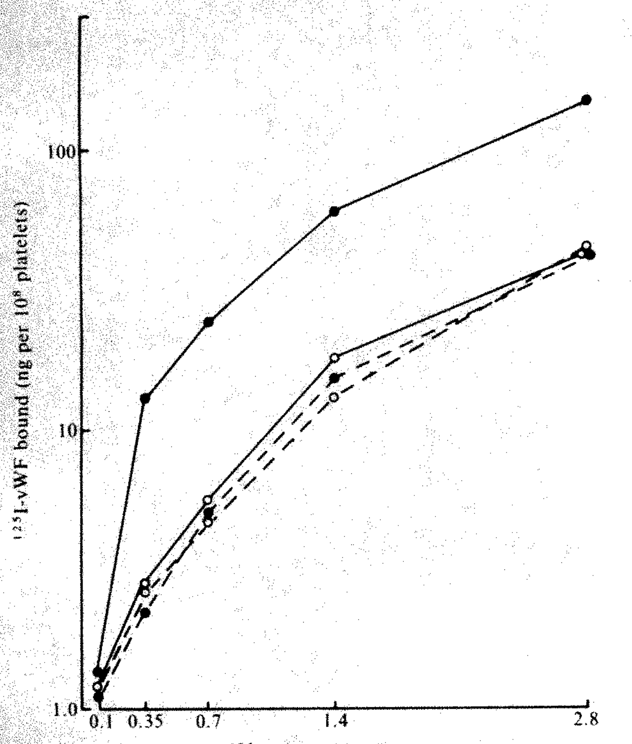


Fig. 1 Binding of ^{125}I-labelled human vWF to platelets treated with 5 μM ADP (●—●); 5 μM ADP plus apyrase (5.4 U; ●- -●), buffer (control: ○—○) and buffer plus apyrase (control: ○- -○). Platelets were treated with apyrase or buffer for 3 min followed by the addition of ADP. After 3 min ^{125}I-vWF was added. Binding was determined after 60 min incubation at room temperature without stirring.

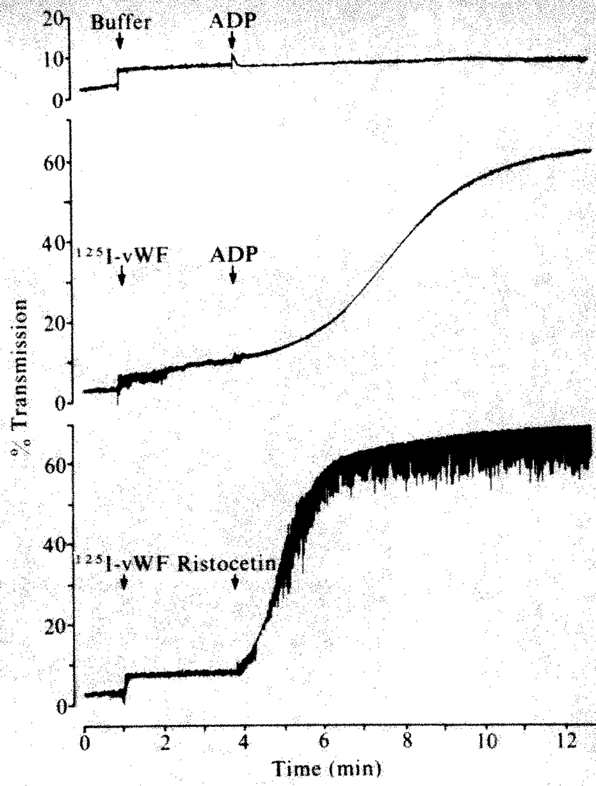


Fig. 2 Aggregation of platelets separated from plasma proteins by ADP and vWF (actual tracings). Upper panel, ADP (5 μM) without vWF; middle panel, ADP+vWF (30 μg); lower panel, ristocetin (1 mg ml^{-1})+vWF.

If binding of multimeric ^{125}I-vWF to the platelet surface is a prerequisite for their aggregation or agglutination, as postulated in the case of ristocetin-induced binding[7], then platelets stimulated with ADP in the presence of ^{125}I-vWF should also aggregate. Indeed, the induction of vWF binding to human platelets by ADP was paralleled by their aggregation. Human platelets, separated from plasma proteins, did not aggregate in response to either ADP or ristocetin[11]. However, addition of purified vWF (30 μg) induced aggregation of both ADP- and ristocetin-treated platelets (Fig. 2). As the fibrinogen content of our purified vWF preparation was <0.00006%, and since platelets were not responsive to ADP in the absence of added ^{125}I-vWF, we believe that the observed aggregation of platelets resulted from the induction of the specific receptor for vWF by ADP and the subsequent linking of platelets by bound multimeric vWF.

The previously observed interference by ADP of human vWF-stimulated platelet agglutination in the presence of EDTA and ristocetin[15] can be attributed to changes in the vWF receptor induced by a combination of ADP, ristocetin and EDTA. It is possible that this combinaton causes a downward shift in the affinity of the receptor for vWF, in contrast to the effect of ADP alone as observed in our system requiring divalent cations.

A major question pertains to nature of the trigger responsible for activation of platelets *in vivo*. Our results support the view of Born[1] that ADP is one such trigger, which by promoting interaction of vWF with platelets allows them to adhere to the vessel wall and to aggregate. *In vivo* activation of platelets by ADP administered intravascularly by microiontophoresis causes them to adhere to the vessel wall[16]. Although platelet activation *in vivo* may occur within 10–100 ms (ref. 17), results

obtained in the inherently slower *in vitro* system can still provide a useful new clue to the *in vivo* situation. According to the mechanism we propose here, as vWF is present at the interface between blood and the vessel wall, activation of platelets by ADP results in exposure of receptors for both vWF and fibrinogen[18-23]. The ADP-induced receptor for vWF binds this multimeric glycoprotein to form an anchor for the attachment of platelets to the inner zone of the vessel wall and to each other.

Bovine thrombin, hirudin, ADP, apyrase, creatinine phosphate and creatinine phosphate kinase were purchased from Sigma. This work was supported by USPHS grants HL-25935, HL-25107 and HL-27560.

Received 25 September 1981; accepted 1 March 1982.

1. Born, G. V. R. in *Platelets and Thrombosis* (eds Sherry, S. & Scriabine, A.) 113–126 (University Park Press, Baltimore, 1974).
2. Tschopp, T. B., Weiss, H. J. & Baumgartner, H. R. *J. lab. clin. Med.* **83**, 296–300 (1974).
3. Zucker, M. B., Kim, S. J., McPherson, J. & Grant, R. A. *Br. J. Haemat.* **35**, 535–549 (1977).
4. Schneider-Trip, M. D., Jenkins, C. S. P., Kahle, L. H., Sturk, A. & ten Cate, J. W. *Br. J. Haemat.* **43**, 99–112 (1979).
5. Jenkins, C. S. P., Clemetson, K. J. & Luscher, E. F. *J. lab. clin. Med.* **93**, 220–231 (1979).
6. Morisato, D. K. & Gralnick, H. R. *Blood* **55**, 9–15 (1980).
7. Kao, K. J., Pizzo, S. V. & McKee, P. A. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5317–5320 (1979).
8. Howard, M. A. & Firkin, B. G. *Thromb. Diath. haemorrh.* **26**, 362–369 (1971).
9. White, J. G. in *Hematology* (eds Williams, W. J., Bentler, E., Erslev, A. J. & Rundles, R. W.) 1023–1039 (McGraw-Hill, New York, 1980).
10. Fujimoto, T., Ohara, S. & Hawiger, J. *J. clin. Invest.* **69** (in the press).
11. Timmons, S. & Hawiger, J. *Thromb. Res.* **12**, 297–306 (1978).
12. Hawiger, J., Niewiarowski, S., Gurewich, V. & Thomas, D. P. *J. lab. clin. Med.* **75**, 93–108 (1970).
13. Miletich, J. P., Jackson, C. M. & Majerus, P. W. *J. biol. Chem.* **253**, 6908–6916 (1965).
14. Born, G. V. R. *Nature* **194**, 927–929 (1962).
15. Grant, R. A., Zucker, M. B. & McPherson, J. *Am. J. Physiol.* **230**, 1406–1410 (1976).
16. Begent, N. A. & Born, G. V. R. *Nature* **227**, 926–930 (1970).
17. Born, G. V. R. *Ciba Fdn Symp.* **71**, 61–78 (1980).
18. Mustard, J. F., Packham, M. A., Kinlough-Rathbone, R. L., Perry, D. W. & Rogoeczi, E. *Blood* **52**, 453–466, 1978.
19. Marguerie, G. A., Plow, E. F. & Edgington, T. S. *J. biol. Chem.* **254**, 5357–5363 (1979).
20. Bennet, J. S. & Vilaire, G. *J. clin. Invest.* **64**, 1393–1401, (1979).
21. Hawiger, J., Parkinson, S. & Timmons, S. *Nature* **283**, 195–197 (1980).
22. Niewiarowski, S., Brudzynski, A. Z., Morinelli, T. A., Brudzynski, T. M. & Stewart, G. J. *J. biol. Chem.* **256**, 917–925 (1981).
23. Peerschke, E. I., Zucker, M. B., Grant, R. A., Egan, J. J. & Johnson, M. M. *Blood* **55**, 841–847 (1980).

Glicentin inhibits gastric acid secretion in the rat

P. Kirkegaard, A. J. Moody, J. J. Holst, F. B. Loud, P. Skov Olsen & J. Christiansen

Department of Surgery D, Glostrup Hospital and Institute of Medical Physiology C, University of Copenhagen, DK-2100 Copenhagen, Denmark
and Novo Research Laboratories, DK-2880 Copenhagen, Denmark

Gut glucagon-like immunoreactants (GLIs) are peptides of enteric origin, which have immunoreactivity in common with pancreatic glucagon. First reported in the intestine by Unger *et al.*[1], GLIs are concentrated in the L-cells of the lower small intestine and the colon[2]. The primary structure of gut GLI-1, glicentin, a highly purified form of the major component of gut GLIs[3], has recently been established. Glicentin contains 69 amino acid residues, has a molecular weight (M_r) of 8,128 (ref. 4) and contains the entire glucagon sequence[4,5], thus establishing it as a member of the hormone family which includes glucagon, secretin, gastric inhibitory polypeptide and vasoactive intestinal polypeptide. Until now there have been insufficient amounts of pure glicentin to test its proposed role in the intestinal phase of the body's response to a meal[6], but there is evidence which suggests that, like other members of the hormone family to which it belongs, glicentin can inhibit gastric acid secretion[7-9]. We confirm here that glicentin inhibits pentagastrin-stimulated acid secretion in the rat.

Female Sprague–Dawley rats (~200 g) were used. The rats were fitted with chronic gastric fistulas according to the method of Lane *et al.*[10] and allowed a 2-week post-operative recovery period. The day before the first experiment, the rats were lightly anaesthetized with ether, and a thin polyethylene catheter was placed in a jugular vein for infusion of glicentin, glucagon or saline. The rats were fasted for 24 h before each experiment, but allowed water *ad libitum*. They were kept in raised mesh-bottom cages to prevent coprophagy.

Ten rats were studied on two separate days. For experimental manipulation, the rats were placed in restraining cages and the stomachs were rinsed out by infusing warm saline through the fistula. After 1 h of spontaneous secretion, pentagastrin (25 μg per kg; Peptavlon, ICI) was given subcutaneously (s.c.). A bolus injection of either 2 ml saline or 3 μg per kg glicentin in 2 ml saline was given intravenously (i.v.) followed by infusion of either 0.9% saline (2 ml h^{-1}) or glicentin (600 ng h^{-1}) in 2 ml saline in random order. Another 8 rats received glicentin (120 ng h^{-1}) and 10 rats received glucagon (1,000 ng h^{-1}). All solutions contained 2% human albumin (Behringswerke). The concentration of glucagon-like immunoreactivity in plasma sampled at the end of the infusion were measured, after ethanol-extraction of the plasma (61%), with a radioimmunoassay specific for the glucagon sequence 6–15, which is also present in the glicentin molecule[5]. Gastric juice was collected continuously and sampled at 30-min intervals. Acid output was determined using an autotitrator (Radiometer ABU 13).

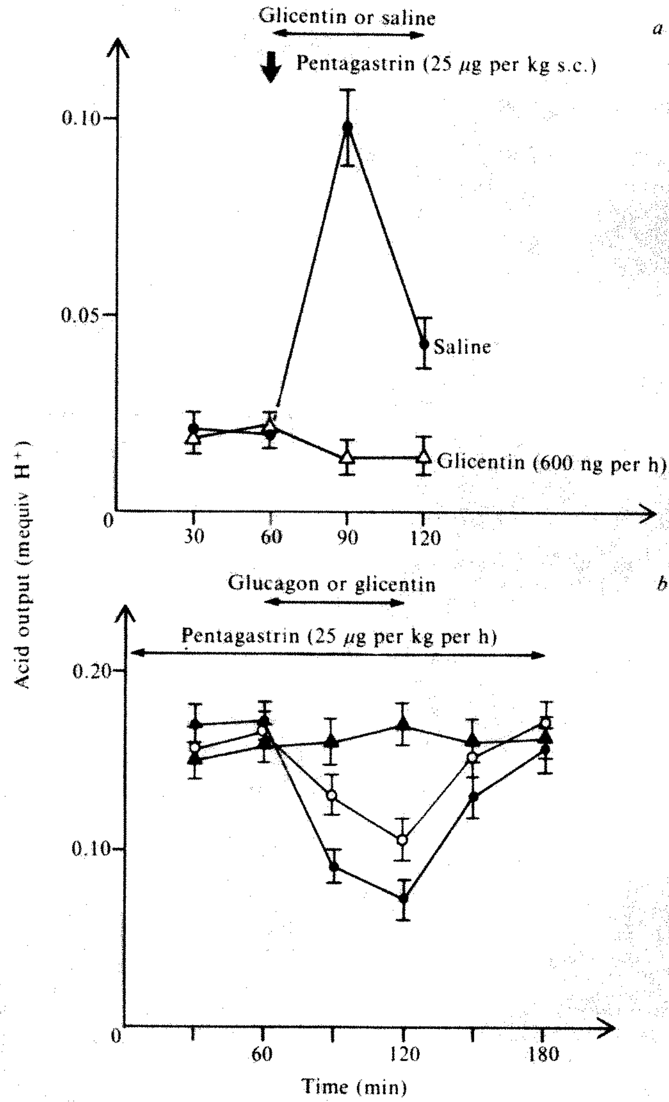


Fig. 1 *a*, Effect of i.v. glicentin infusion on submaximally stimulated gastric acid secretion in the chronic-fistula rat. Acid output (milliequivalents H^+) did not exceed baseline values. Control rats were treated with saline. *b*, Effect of i.v. infusion of glicentin (○, 120 ng h^{-1}; ●, 600 ng h^{-1}) and glucagon (▲, 1,000 ng h^{-1}) on maximally stimulated gastric acid secretion in the chronic-fistula rat. During infusion of glicentin (both doses), acid output decreased significantly. Glucagon had no effect on acid secretion. For *a* and *b*, each point represents 30 min acid secretion (mean ± s.e.m. of 10 rats).

0028-0836/82/190156—02$01.00

In another 10 chronic-fistula rats, pentagastrin (25 μg per kg per h) was infused i.v. for 3 h. After 1 h, a bolus injection of 0.6 or 3 μg per kg glicentin was given and for the next hour, 120 or 600 ng glicentin was infused together with the pentagastrin. For the last hour, pentagastrin alone was infused. The same experiment was performed with glucagon (1,000 ng per h) after a bolus injection of 5 μg per kg. Gastric juice was collected and acid output determined as described above.

The porcine glicentin used was isolated as previously described[3] and gave a single band on analytical isoelectrifocusing and polyacrylamide gel electrophoresis. The content of gastric inhibitory polypeptide was <100 p.p.m. (ref. 11) and that of vasoactive intestinal polypeptide was 70 p.p.m. (Moody, A. J. M., unpublished data) as determined by radioimmunoassay.

We observed a marked effect of glicentin on gastric acid secretion. After submaximal stimulation (pentagastrin, 25 μg per kg s.c.) the acid response was completely abolished by the infusion of glicentin (Fig. 1*a*). After maximal stimulation (continuous i.v. infusion of pentagastrin, 25 μg per kg per h) the acid output decreased in a dose-dependent manner during simultaneous infusion of glicentin. When infusion of glicentin was stopped, there was an immediate increase in gastric acid secretion. Glucagon at 1,000 ng per h had no effect on acid secretion (Fig. 1*b*). The levels of glucagon-like immunoreactivity in plasma obtained after infusion of glicentin at 600 ng per h and glucagon at 1,000 ng per h, were 541±135 and 428±94 pmol-equivalents l^{-1}, respectively, compared with 72±27 pmol l^{-1} in rats receiving saline alone (mean±s.e.m.).

The present study demonstrates that glicentin is a potent acid secretion inhibitor and the fact that it is effective in low doses supports the view that GLI-1 may act as an enterogastrone in the intestinal phase of gastric acid secretion. As glucagon, given in doses greater than those of glicentin and giving rise to similar concentrations of glucagon-like immunoreactivity, was ineffective, it is unlikely that the action of glicentin is due to that part of its sequence which corresponds to the glucagon molecule.

Received 1 December 1981; accepted 18 March 1982.

1. Unger, R. H., Eisenstraut, A. M., Sims, K., McCall, M. S. & Madison, L. L. *Clin. Res.* **9**, 53–60 (1961).
2. Holst, J. J. *Digestion* **17**, 168–190 (1978).
3. Sundby, F., Jacobsen, H. & Moody, A. J. *Hormone Metab. Res.* **8**, 366–371 (1976).
4. Thim, K. & Moody, A. J. *Regul. Peptides* **2**, 139–151 (1981).
5. Holst, J. J. *Biochem. J.* **187**, 337–343 (1980).
6. Holst, J. J. in *Gut Hormones* (ed. Bloom, S. R.) 383–386 (Churchill-Livingstone, Edinburgh, 1978).
7. Christiansen, J. *et al. Scand. J. Gastroenterol.* **14**, 161–166 (1979).
8. D Sa Barro, A. A. B. & Buchanan, K. D. *Gut* **18**, 877–881 (1977).
9. Coudray, A. M. & Rosselin, G. *Biosci. Rep.* **1**, 151–155 (1981).
10. Lane, A., Ivy, A. C. & Ivy, E. *Am. J. Physiol.* **192**, 221–228 (1957).
11. Lauritsen, K. B. & Moody, A. J. *Diabetologia* **14**, 149–153 (1978).

A sequence of mitochondrial DNA is associated with the onset of senescence in a fungus

Corinne Vierny, Anne-Marie Keller, Odile Begel & Léon Belcour

Centre de Génétique Moléculaire, Central National de la Recherche Scientifique, 91190 Gif-sur-Yvette, France

In *Podospora anserina*, a strict aerobic fungus, senescence[1] (arrest of vegetative growth of the mycelium) is correlated with severe modifications of the mitochondrial DNA similar to those observed in the *rho*⁻ mutants of yeast[2]. Senescent cultures yield a specific DNA (SEN-DNA) consisting of small circular molecules arranged in multimeric series of sizes[3,4] and which results from the amplification of either one of two non-overlapping segments of the mitochondrial DNA[5,6]. One region 0.89 μm long [2.6 kilobases (kb)] is most frequently amplified, in different strains of *P. anserina* and also in subcultures of a strain reaching senescence independently[6]. In 'young' cultures of the mating type-minus wild-type strain s (*s mt*⁻), the 2.6 kb region exists both as an integral part of the mitochondrial chromosome and as rare free sequences[6]. The results presented here show that (1) the free form is absent from two strains, *s mt*⁺ and *s mt*⁻ *cap*ʳ*1*, which both display increased longevity and which differ respectively from the *s mt*⁻ strain by a nuclear allele at the *mt* locus and by a mitochondrial mutation conferring resistance to chloramphenicol; and (2) in the mutant *mex1*, which develops beyond the edge of arrest of a senescent mycelium, that is, escapes senescence, the mitochondrial DNA precisely lacks the 2.6 kb region. The results suggest that excision and/or amplification of the 2.6 kb sequence is the cause of senescence.

We previously reported that senescent cultures of *P. anserina* contain one of two types of SEN-DNA differing by their density[6]. The first type has a density of 1.694 g cm^{3} (identical to that of the normal 'young' mitochondrial DNA) and exhibits a variable monomer unit comprising 4–6.3 kb. Cross-hybridization revealed that all SEN-DNAs of density 1.694 g cm^{-3} so far obtained share a sequence of mitochondrial DNA of ~2.3 kb and thus belong to the same family. The second type of SEN-DNA has a density of 1.699 g cm^{-3}, a size of 2.6 kb and a sequence that has been reproducibly recovered from independent senescent cultures grown in this and other laboratories[4,6,7]. Hybridization of SEN-DNAs with the mitochondrial DNA of a young culture of the *s mt*⁻ strain digested by *Hae*III (see Table 1 for a description of the strain) shows that both types of SEN-DNAs result from the amplification of particular regions of the mitochondrial chromosome[5,6]. However, whereas the 1.694 g cm^{-3} SEN-DNA hybridizes only with fragments of the 'young' mitochondrial DNA visible on the gel, the 1.699 g cm^{-3} SEN-DNA hybridizes to a 2.6 kb position as well, at which no visible fragment is found (see Fig. 1*a* and its legend). We have therefore assumed that the 2.6 kb sequence is present in young *s mt*⁻ cultures not only as an integral part of the mitochondrial chromosome but also as rare free molecules[6]. Three properties of the 1.699 g cm^{-3} SEN-DNA—its specific density, its peculiar pattern of hybridization with the 'young' *s mt*⁻ mitochondrial DNA and its repeated appearance in senescent cultures of independent origin—distinguish it from 1.694 g cm^{-3} SEN-DNAs and suggest that it has episome-like properties. In any case, whatever the basis for the difference between the two types of sequences, it is clear that senescence is always correlated with the amplification of one of these types of sequence. This correlation raises the question of the cause-effect relationship between senescence and the presence of the amplified sequences, and it is tempting to assume that the circular amplified molecules constituting the SEN-DNAs correspond to the infectious cytoplasmic factor responsible for senescence and identified by Marcou[8].

Two sets of new data on the 2.6 kb sequence provide good support for the idea that senescence actually results, at least in some cases, from the amplification of this sequence. (1) The *Hae*III-digested mitochondrial DNAs extracted from young cultures of the two strains *s mt*⁺ and *s mt*⁻ *cap*ʳ*1* (whose life spans are longer than that of strain *s mt*⁻; Table 1) were hybridized with the 1.699 g cm^{-3} SEN-DNA. In both cases, no

Table 1 Life spans of the different strains used

Genotype: Nuclear	Genotype: Mitochondrial	Life span (cm)
s mt⁻	*cap*ˢ	14
s mt⁺*	*cap*ˢ	40
s mt⁻†	*cap*ʳ*1*	100
s mt⁻	*cap*ˢ *mex1*	>110

* Strain *s mt*⁺ is isogeneic with strain *s mt*⁻ except for the mating-type locus and the two strains are cytoplasmically identical.

† Strain *s mt*⁻ *cap*ʳ*1* is strictly isogeneic with strain *s mt*⁻ from which it derives, except for the mitochondrial mutation *cap*ʳ*1*.

0028–0836/82/190157—03$01.00

labelled material was recovered at the 2.6 kb position (Fig. 1*A*, parts *b* and *c*). This indicates that in the strains *s mt*$^+$ and *s mt*$^-$ *cap'1*, the 2.6 kb sequence does not exist in a free form but is present only as an integrated sequence. (2) A mutant *mex1*, having a modified lifespan, was obtained and studied. This mutant was selected as a sector of growing mycelium that developed beyond the arrested edge of a *s mt*$^-$ senescent culture. Such revertants of the senescent state are uncommon. The mutant *mex1* exhibits a slow growth rate but since its isolation it has been growing steadily and has shown no symptoms of senescence. Preliminary genetic analysis of mutant *mex1* indicated that it segregates in a non-mendelian way, which suggested that the mutation might be located in the mitochondrial DNA. Restriction analysis of the mitochondrial DNA of this mutant confirmed this hypothesis. Figure 2 shows the restriction pattern (after digestion by *Bgl*II) of the mitochondrial DNA of mutant *mex1* compared with that of the wild-type strain

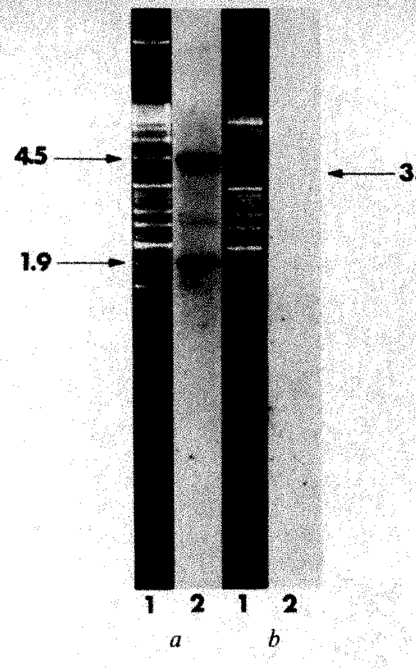
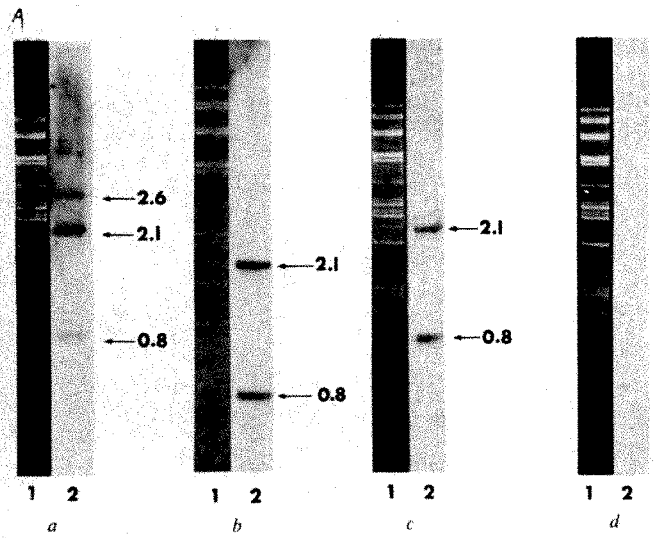


Fig. 2 Lanes 1, electrophoretic patterns of mitochondrial DNA of strains *s mt*$^-$ (*a*) and *s mt*$^-$ *mex1* (*b*) digested by *Bgl*II. Lanes 2, autoradiographs after transfer of the fragments to nitrocellulose filter and hybridization with a ^{32}P-labelled 1.699 g cm^{-3} SEN-DNA obtained by nick-translation (*a*, strain *s mt*$^-$; *b*, strain *s mt*$^-$ *mex1*). Note the monomer unit of the 2.6 kb SEN-DNA possesses one site for the restriction enzyme *Bgl*II. Sizes in kb.



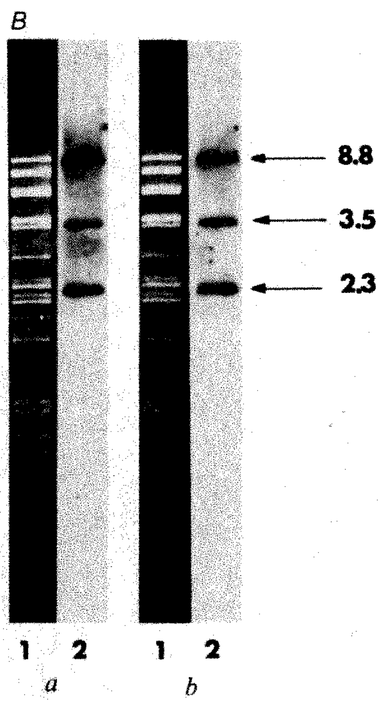


Fig. 1 *A*, Southern hybridizations between *Hae*III-digested 'young' mitochondrial DNAs of different strains and ^{32}P-labelled 1.699 g cm^{-3} nick-translated SEN-DNA. The conditions for preparation and endonuclease digestion of DNA, gel electrophoresis and transfer of DNA to nitrocellulose filters were as previously described[5]. The labelling of DNA by nick-translation was performed as described in ref. 12. For each DNA studied, the electrophoretic pattern and the autoradiograph of the filter after hybridization are shown in lanes 1 and 2 respectively. Sizes in kb. *a*, Strain *s mt*$^-$; *b*, strain *s mt*$^+$; *c*, strain *s mt*$^-$ *cap'1*; *d*, strain *s mt*$^-$ *mex1*. Note that the molecules which constitute the 2.6 kb SEN-DNA are circular and that the monomer unit of this SEN-DNA has a single *Hae*III site. Only one fragment of 2.6 kb is thus recovered after digestion of this SEN-DNA by *Hae*III. *B*, Southern hybridization between *Hae*III-digested 'young' mitochondrial DNAa of strains *s mt*$^-$ (*a*) and *mex1* (*b*) and ^{32}P-labelled 1.694 g cm^{-3} nick-translated SEN-DNA (monomer unit = 6.3 kb). In each case, the electrophoretic pattern and the autoradiograph of the filter after hybridization are shown in lanes 1 and 2 respectively. Sizes in kb. Note that the monomer unit of the 6.3 kb SEN-DNA possesses two *Hae*III sites.

from which it derives. The *mex1* mitochondrial DNA lacks two bands of respectively 4.5 and 1.9 kb and presents an additional band of 3.8 kb, thus *mex1* has a deletion of ~2.6 kb. Interestingly, the missing fragments are precisely those previously shown to hybridize with the 1.699 g cm^{-3} SEN-DNA. The number and the nature of the fragments missing in *mex1* and the estimated size of the deleted fragment obviously suggested that the mutation *mex1* corresponded to the deletion of the 2.6 kb sequence which is often amplified in senescent cultures. This was confirmed by hybridization experiments on the mitochondrial DNA of both young wild-type and *mex1* cultures after digestion by *Bgl*II, using a 1,699 g cm^{-3} nick-translated SEN-DNA as ^{32}P-labelled radioactive probe. The results of this experiment (Fig. 2) show that the mitochondrial DNA of mutant *mex1* does not exhibit any homology with the 1.699 g cm^{-3} SEN-DNA. Similar hybridization experiments in which the mitochondrial DNA *mex1* is digested by *Hae*III (Fig. 1*A*, *d*) or *Bam*HI yield the same result. It can be concluded that in the mutant *mex1*, a sequence containing at least the monomer of the 1.699 g cm^{-3} SEN-DNA has been deleted. The electrophoretic pattern of the *Hae*III- (or *Bgl*II-) digested mitochondrial DNA of mutant *mex1* indicated the absence of any modifications in the region capable of giving rise to the 1.694 g cm^{-3} family of SEN-DNAs. This was confirmed by experiments in which *Hae*III-digested *mex1* and *s mt*$^-$ mitochondrial DNAs were hybridized with 1.694 g cm^{-3} ^{32}P-labelled nick-translated SEN-DNA (monomer unit = 6.3 kb); the same hybridization pattern was obtained in both cases (Fig. 1*B*).

The data presented above show clearly that the 2.6 kb sequence has a central, although perhaps not unique, role in the determination of senescence. The existence of a mutant missing this sequence, which after more than 1 m of growth (see mean life span of the corresponding wild-type strain in Table 1) had not yet become senescent, suggests that excision and/or amplification of the 2.6 kb sequence are the cause of senescence or at least necessary for its onset. This hypothesis is supported by hybridization experiments which reveal that the state in which this sequence exists in young cultures (free and integrated in strain *s mt*$^-$, integrated only in strains *s mt*$^+$

and *s mt⁻ cap'1* and absent in strain *mex1*) seems correlated with the longevities of the strains; the higher the probability of appearance of the free 2.6 kb sequence, the shorter the longevity of the strain. The fact that the strains *s mt⁺* and *s mt⁻ cap'1* are isogeneic except for a nuclear (mt^+/mt^-) or a mitochondrial (cap^r1/cap^s1) gene implies that these might act on, for example, the probability of excision or the efficiency of amplification of the 2.6 kb sequence. Nothing is yet known about the mechanisms responsible for the appearance of the 1.699 g cm^{-3} SEN-DNA. However, the presence of a population of molecules lacking the 2.6 kb sequence in the residual mitochondrial DNA of senescent cultures displaying this amplified sequence[9] (our results, in preparation) suggests that excision rather than replication is the primary event necessary for further amplication of this sequence. The existence of a mutant lacking the 2.6 kb sequence indicates that the sequence does not code for a product essential for cell survival. The reduced growth rate of mutant *mex1* suggests, however, that the sequence is physiologically important, perhaps, as already postulated[6], in initiating the replication of the mitochondrial chromosome. As more than one replication origin is likely, as demonstrated in yeast[10,11], the mitochondrial chromosome would still be able to replicate in the absence of the 2.6 kb sequence but at a slower rate.

The phenomenon of amplification during senescence is not exclusive to the 2.6 kb sequence. Another family of sequences sharing a common region can also be amplified and, depending on the senescent culture, this second type of sequence is amplified either alone[5,6] or in association with the 2.6 kb sequence (our unpublished results). In the latter case, it is not yet clear whether the amplification of the two types of sequence is related, although preliminary results of other senescence-resistant mutants suggest that the absence or rearrangement of the second family of sequences also prevents the appearance of senescence.

This work was supported in part by INSERM grants 794 0062 and 0038279114. We thank Dr Janine Beisson for her careful and critical reading of the manuscript.

Received 19 January; accepted 17 March 1982.

1. Rizet, G. *C.r. hebd. Séanc. Acad. Sci., Paris* **237**, 838–840 (1953).
2. Slonimski, P. P. & Lazowska, J. *Mitochondria 77* (ed. Bandlow, W.) 39 (De Gruyter, Berlin, 1977).
3. Stahl, U., Lemke, P. A., Tudzynski, P., Kück, U. & Esser, K. *Molec. gen. Genet.* **162**, 341–343 (1978).
4. Cummings, D. J., Belcour, L. & Grandchamp, C. *Molec. gen. Genet.* **171**, 239–250 (1979).
5. Jamet-Vierny, C., Begel, O. & Belcour, L. *Cell* **21**, 189–194 (1980).
6. Belcour, L., Begel, O., Mossé, M. O. & Vierny, C. *Curr. Genet.* **3**, 13–21 (1981).
7. Stahl, U., Kück, U., Tudzynski, P. & Esser, K. *Molec. gen. Genet.* **178**, 639–646 (1980).
8. Marcou, D. *Annls Sci. nat. bot.* **11**, 653–764 (1961).
9. Kück, U., Stahl, U. & Esser, K. *Curr. Genet.* **3**, 151–156 (1981).
10. Blanc, H. & Dujon, B. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3942–3946 (1980).
11. Gourshot, R., De Zamaroczy, M., Baldacci, G. & Bernardi, G. *Curr. Genet.* **1**, 173–176 (1980).
12. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. *J. molec. Biol.* **113**, 237–251 (1977).

Structural analysis of insertion sequence IS*5*

Manfred Kröger & Gerd Hobom

Institut für Biologie III der Universität Freiburg, Schänzlestrasse 1, D-7800 Freiburg i.Br., FRG

Among the transposable DNA elements that have been detected in bacterial genomes and plasmids[1–3], the small insertion sequence (IS) elements have been recognized as a class of autonomous, basic units of transposition. IS elements are not only fully competent of transposition and able to cause several other related chromosomal aberrations such as deletion of adjacent DNA segments[4,5] and integration of circular plasmid DNAs[6], but have also proved to be the active units of transposition in the larger composite transposon (Tn) elements[1–3,7–9]. In addition, IS and Tn elements interfere with local transcription patterns in the host genome by causing polarity [10,11]. Structural and functional analysis of the IS elements should therefore elucidate both the basis for autonomy of a genetic unit as small as 1,000 base pairs (bp), and the transposition mechanism, including its inherent step of DNA duplication[12]. We have now determined the 1,195 bp sequence of the transposable DNA element IS*5* in phage λKH100. Near-terminal signal structures appear to protect the central region from outside-in transcriptions and translations, and render IS polarity as a direct consequence of IS autonomy. The nucleotide structure shows an unusual, highly compact organization of two completely overlapping, antiparallel genes and correlated control sequences.

IS*5* was first identified among clear plaque mutants of bacteriophage λ (λKH100), and was accordingly localized to the C-terminus of the *c*I repressor gene[13,14]. IS*5* was estimated to be 1,200–1,250 bp long, and electron microscopy revealed short inverted sequence repetitions at its ends. The IS*5* unit of λKH100 seemed different from IS*1*–IS*4*, but identical with two other insertions respectively in the bacterial *arg*B gene of specialized transducing phage λ13d*arg*B2 and in phage Mu DNA[15]. Both ends of IS*5* are known to be deletogenic[15] and to stimulate general recombination within adjacent DNA segments[16]. IS*5* also appears to exert polar effects in at least one orientation[17,18].

We have used λKH100 DNA as a primary source of IS*5*, and have constructed a hybrid plasmid pHL118 by cloning a 1,720 bp *Hin*dIII fragment into a pBR313-derived vector plasmid (see ref. 12). The λKH100 *Hin*dIII fragment contains all of the IS*5* DNA as well as 564 bp of λDNA (λ: 37,727–38,291), and it was used for DNA sequencing according to the Maxam and Gilbert method[19] (see Fig. 1). Mapping of the resulting restriction fragments has provided independent proof of the existence and location of a large number of endonucleolytic recognition sites (see Fig. 2). Both junctions and the λKH100 DNA sequences adjacent to IS*5* have also been sequenced and found to be consistent with published data[20,21] obtained from other λ phage or plasmid DNAs. IS*5* is only slightly smaller than was estimated by electron microscopy[15], and the position of IS*5* in the λKH100 sequence (codon 207 of the 232 codon *c*I gene, see Fig. 2*a*) is only slightly to the left of two genetic determinations[13,16].

When IS*5* is inserted into λ*c*I DNA, 4 bp of the wild-type sequence (target site TTAG; λ:38,150–38,153) are directly duplicated to the left and right of both IS*5* ends, and are connected to the two IS*5* terminal inverted repeat sequences of 16 bp (with a single mismatch at position 13) (see Fig. 2*a*). No internal repetition of the inverted repeat sequence has been detected. It is now clear that a partial sequence of phage Mu445-8 DNA[22] includes an identical 99 bp fragment from the right *Eco*RI end of IS*5* (with a single deviation at IS*5*:1,171), followed by an identical target site sequence in opposite orientation (CTAA). A 4 bp target site duplication has not previously been observed for the integration of prokaryotic transposable elements, but has been reported in the integration of Moloney strains of murine leukaemia and sarcoma viruses in eukaryotic DNA[23,24].

Most of the internal sequence of IS*5* is dominated by a very long open reading frame, which extends in right to left direction from position IS*5*:1,127 to IS*5*:147 and codes for a 326 amino acid, predominantly basic protein of molecular weight $37{,}831^{+60}_{-38}$. The IS*5* large gene coding region is preceded by a Shine–Dalgarno[25] initiation sequence (the most likely fit is shown in Fig. 2*f*). A computer analysis also identified potential promoter and terminator signals[26] before (IS*5*:1,175–1,139)

0028–0836/82/190159—04$01.00

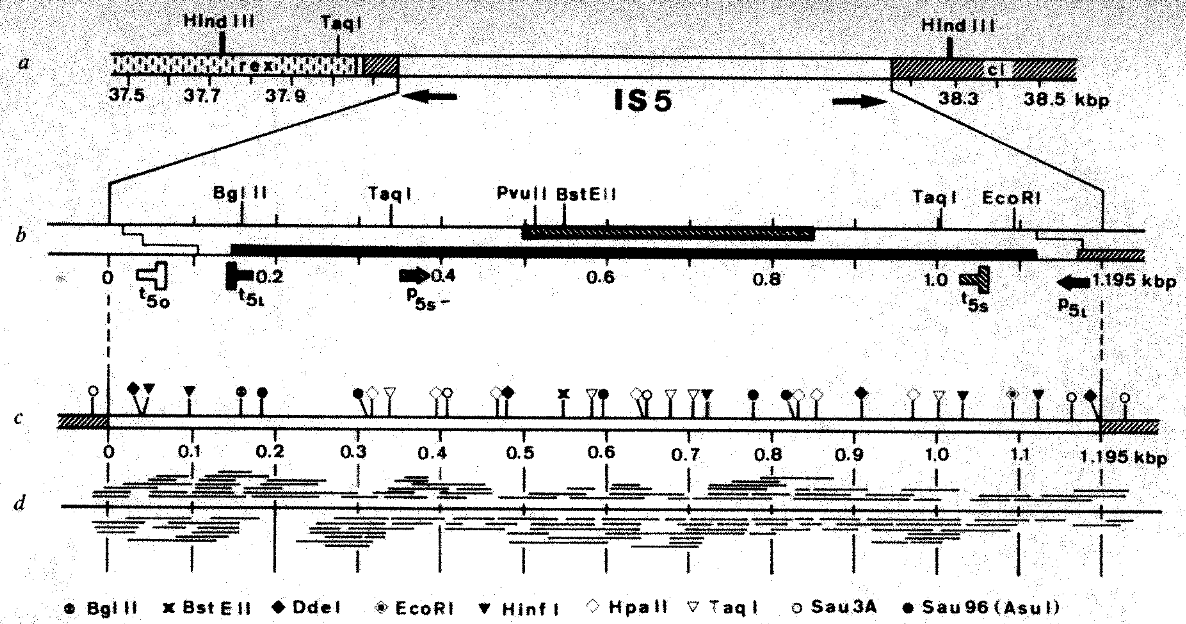


Fig. 1 Restriction map of the *rex-c*I region of λKH100 DNA and sequencing chart for insertion element IS*5* DNA. *a*, Position of insertion IS*5* in the *c*I gene of λKH100. Hatched regions represent IS*5*-adjacent λ DNA; sequence positions in λ DNA refer to the *Eco*RI restriction site located at the origin of replication (position 40,000). The marked *Hin*dIII fragment was cloned into a pBR313-derived vector plasmid and propagated as described previously[33]. *b*, Genetic map of insertion element IS*5* from the complete nucleotide sequence in Fig. 2 and from a functional analysis of its terminator and promoter signals of transcription (see Fig. 3 and text). The black bar indicates the coding region for the large IS*5* protein, the hatched bar corresponds to the IS*5* small protein reading frame. Outside-in translation according to the diverted λ*c*I reading frame is indicated as a lightly hatched bar at the right IS*5* boundary, and the near-terminal stop codons for the other five reading frames are indicated at the left and right termini respectively. Three transcriptional terminator signals (t_{5o}, t_{5l}, t_{5s}) and two promoter sequences (p_{5l}, p_{5s}) are indicated in their tentative locations. The restriction sites used during functional studies (see Fig. 3 legend) are shown above the genetic map. *c*, Restriction map of IS*5* DNA, indicating the restriction sites used in generating 5'-labelled fragments for DNA sequencing: *Bgl*II, *Bst*EII, *Dde*I, *Eco*RI, *Hin*fI, *Hpa*II, *Sau*3A, *Sau*96 and *Taq*I. A key for restriction site symbols is given at the bottom of the figure. *d*, Summary of DNA sequence results obtained for IS*5* DNA. Every stretch of independently determined DNA sequence is indicated by a single horizontal bar covering the corresponding region of unequivocal reading; segments of *l*-strand DNA are given above, *r*-strand segments below the central line. Note that almost the entire element has been sequenced using both strands, and large sections have been sequenced repeatedly, from different endonuclease cuts, as indicated by piles of horizontal bars. DNA sequencing was by the Maxam–Gilbert method, using the A > C reaction to determine adenine positions[19]. All isolated fragments were purified by butanol extraction[34] before both the kinase reaction and chemical degradation steps.

and after (IS*5*:166–143) the large gene coding sequence respectively.

Although none of the other two reading frames in right to left direction is open for any considerable length, the IS*5* DNA sequence appears to have the coding potential for a second protein in left to right direction, from position IS*5*:525 to IS*5*:851. The second open reading frame, which is in register with the sequence coding for the larger IS*5* protein, is predicted to code for a 108 amino acid, predominantly apolar protein of molecular weight $11{,}069^{+11}_{-10}$. Analogous to the IS*5* large gene, the IS*5* small gene also is preceded by a ribosome recognition sequence (see Fig. 2*g*; an alternative AUG start codon in the same reading frame, at IS*5*:495, is not affiliated with a recognizable Shine–Dalgarno sequence), and , according to computer search, has correlated promoter and terminator signals (at IS*5*:355–401 and 1,019–1,054) before and after the coding sequence.

To measure IS*5* promoter and terminator activities *in vivo*, we and Rak *et al.*[27] have individually subcloned into hybrid plasmids certain fragments containing IS*5* signal elements. Our analysis of the IS*5* terminator fragments (Fig. 3) confirmed computer-predicted locations and orientations, within the present limits of minimization. In addition to the large gene and small gene terminators (t_{5l} and t_{5s}), a termination signal has been detected near the left terminus of IS*5* in outside-in orientation (t_{5o}; IS*5*:31–71) which is consistent with the observed polarity of the element[18]. No analogous terminator has been found near its right end, however. The position of the IS*5* large gene promoter (p_{5l}) has been confirmed by Rak *et al.*, while the IS*5* small gene promoter (p_{5s}) was located at IS*5*: 460–495 in their cloning analysis. Expression of IS*5* large and small proteins has also been detected in minicells[27].

No other example of the IS*5* type of DNA dual coding capacity has been reported. Computer screening of published IS DNA sequences—IS*2* (ref. 28), IS*4* (ref. 29) and IS*903* (ref. 30) has, however, revealed quite similar reading frame organizations. In addition to a dominating large reading frame of more than 300 codons, there is always a second reading frame of 108–143 codons in the opposite orientation. Note that the large and small reading frames are consistently found in register, and that the (presumptive) small gene is always completely covered by the large gene. The sequence of IS*1*[31,32] shows a similar pattern, although both the large and small reading frames are short: 167 and 89 codons respectively. Similar to IS*5* (Fig. 2*g*), the small gene Shine–Dalgarno recognition sequences always appear as part of mRNA hairpin structure covering the smaller reading frame initiation region, and may be indicative of a generally reduced expression. There is no evidence, however, in any other element of promoter or terminator transcription signals or of dual protein expression. The role of such a compact antiparallel arrangement in genetic regulation or other functions is unknown.

An IS element preserves its autonomy by a genetic organization which avoids regularity dominance by adjacent genes or transcription units on the host DNA. There are therefore near-terminal stop codons in all six reading frames (marked in Fig. 2*b*) to avoid formation of fusion proteins under foreign control. In addition, near terminal termination signals of transcription, and therefore polarity, may be a direct consequence of IS autonomy. Whereas the structure of IS*4*, which has two outside-in termination signals[29], fully agrees with this prediction, IS*5* does not carry a termination signal at its functionally rather crowded right-hand end. IS autonomy may be protected in other ways, however (compare Fig. 2*f*). The biological significance of the different organization of the two ends of IS*5* is not known.

We thank B. Rak for communicating and discussing results before publication, and B. Traub and G. Schlingmann for technical assistance. λKH100 was kindly provided by W. Szybalski.

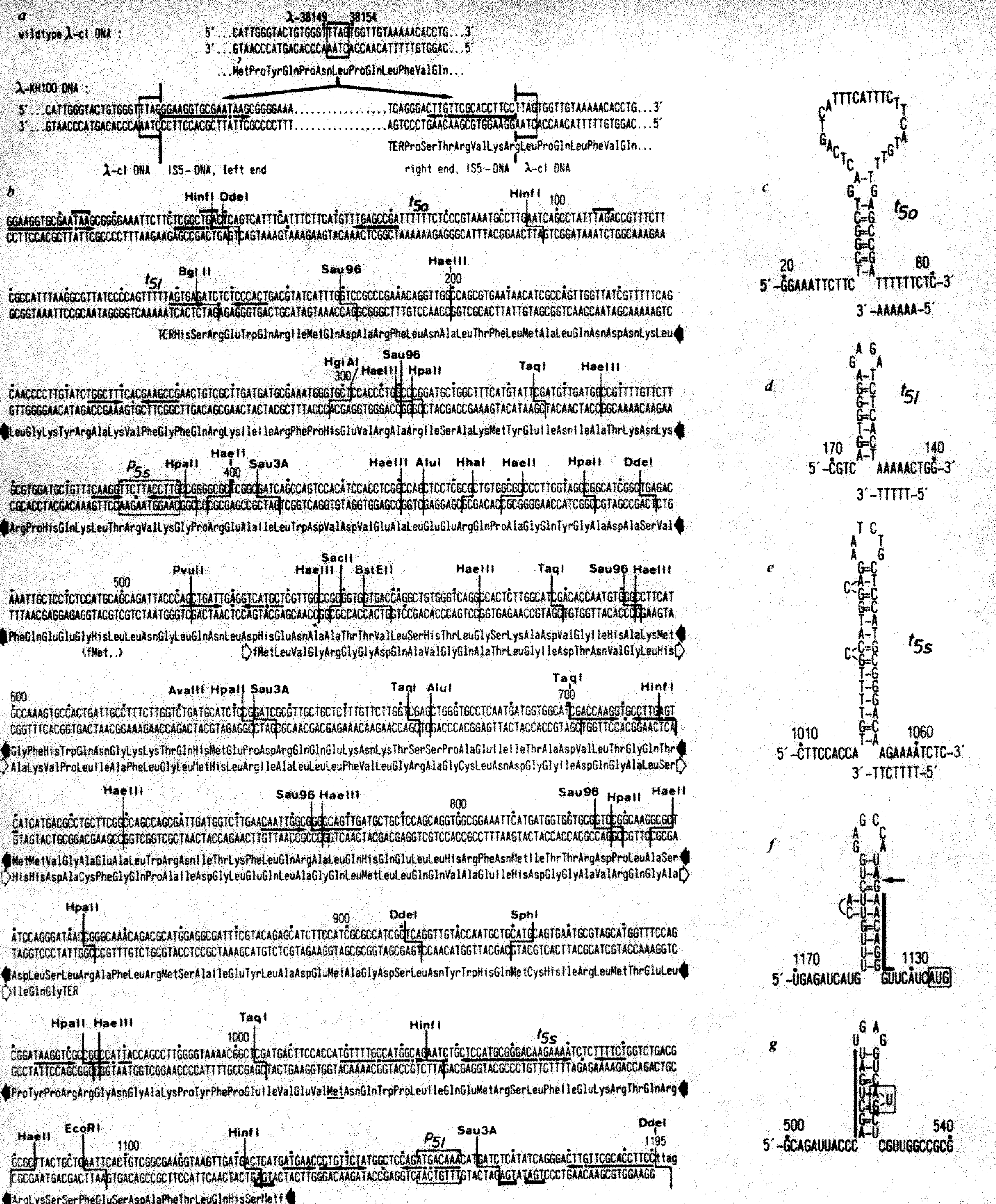


Fig. 2 DNA sequence of insertion element IS*5* in λKH100 DNA. *a*, Identification of the IS*5* integration site in the coding sequence of the *c*I repressor gene of phage λKH100. The 4 bp target site duplication is boxed, and the 16 bp terminal inverted repeat sequences (with a single mismatch) indicated by arrows. A proportion of the λ wild-type and of λKH100 *c*I protein sequences are shown below the DNA sequences. According to the genetic map of phage λ, the termini of IS*5* have been named left and right end respectively. *b*, Complete nucleotide sequence of IS*5*. Numbering starts from the left end. Amino acid sequences predicted from the DNA sequences of the large (right to left; closed arrows) and small (left to right; open arrows) genes of IS*5* are given in separate lines below the DNA sequence. Two potential alternative initiation sites for the large and small genes are indicated at positions 1,019 and 495 respectively. Cleavage sites for a collection of restriction enzymes that have been used for fragment mapping or DNA sequencing are marked at the respective nucleotide position above the DNA sequence. (In addition to experimentally proved restriction cuts, the recognition sequences for *Ava*III, *Hae*II and *Sph*I are also indicated.) Arrows between the DNA strands designate inverted sequence repetitions. Brackets refer to near-terminal stop codons with only the one closest to the terminus indicated for each of the six reading frames (excluding those in the target site duplication). Both rightward and leftward promoters are indicated by boxes at their tentative locations. Methylated cytidine *Eco*RII/*Bst*NI recognition sequences are indicated by squares; cleavage at the *Sau*96/*Hae*III restriction site at IS*5*:310 occurred in low yields, apparently due to an overlap with a methylated C residue. *c–e*, DNA secondary structures predicted for transcription terminator signals of IS*5*. Terminators for rightward outside-in transcription, for the IS*5* large gene (leftwards) and the IS*5* small gene transcription (rightwards), respectively, are drawn in a non-coding strand hairpin/coding strand 'tail' sequence manner (see refs 33, 35). Numbers refer to the respective positions in the IS*5* DNA sequence. *f*, *g*, RNA secondary structures predicted to occur at the initiation regions for the IS*5* large and small gene proteins respectively. The RNA hairpin structure in *f* can only be formed in a (long) outside-in transcription messenger and not within a (short) mRNA started at p_{5l}. According to a best fit with a bacterial promoter consensus sequence [26], the 5′ end of the p_{5l} mRNA is located at the position marked by an arrow (IS*5*:1,142), which would preclude formation of an initiation region hairpin structure. AUG initiator codons are boxed; bars along the sequences refer to Shine–Dalgarno recognition sequences[25].

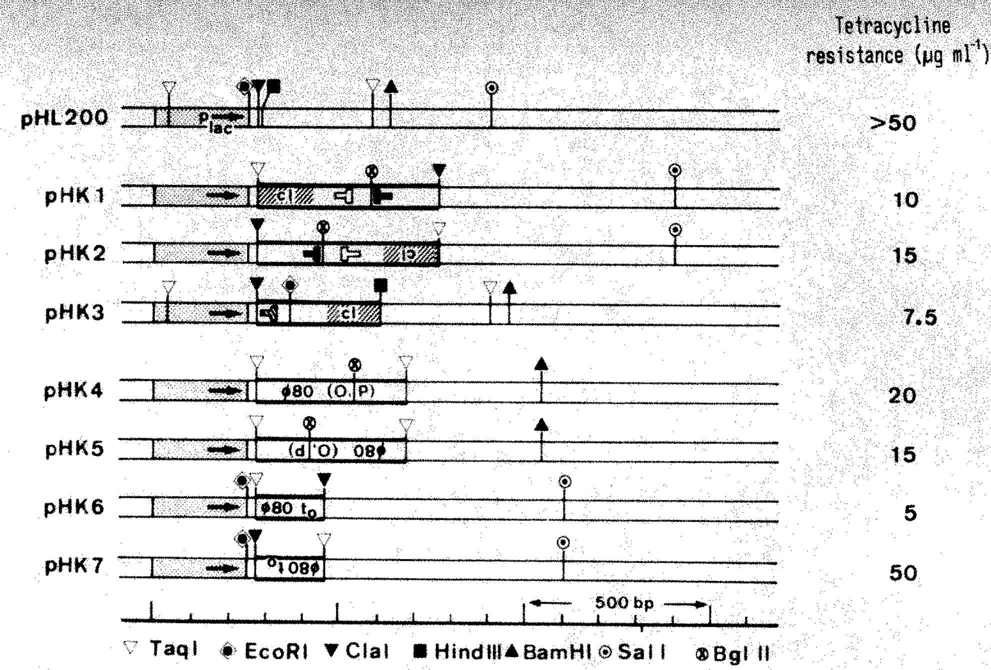


Fig. 3 Hybrid plasmid functional analysis of IS*5* terminator fragments. For all cloning experiments the same vector DNA, pHL200, has been used. This vector is a derivative of pBR322 and carries to the left of the *Eco*RI site a *lac*UV5 promoter fragment ~286 bp long (stippled area with arrow pointing in direction of RNA synthesis). The starting material to construct pHL200, pEx(lac)UV5-150, was kindly provided by H. Weiher, B. Zink and H. Schaller. Insertion of foreign DNA fragments between the *lac* promoter and downstream tetracycline-resistance genes made use of the *Cla*I and *Hin*dIII cloning sites. The tetracycline promoter of pBR322 extends across these restriction sites, but is destroyed by insertion. Care had to be taken, however, not to reconstitute the tetracycline promoter sequence through insertion of sequences similar to the missing −35 region. The levels of tetracycline resistance have been measured on the same freshly prepared agar plates for the whole set of clones together with a pBR322-containing strain (>50 μg ml^{-1} tetracycline) and an 'empty' strain of *Escherichia coli* K-12−490A (3 μg ml^{-1} tetracycline) as controls. The respective limiting concentrations of tetracycline are given on the right. DNA fragments incorporated into pHL200 are shown in heavier lines. The cleavage sites of restriction enzymes used for cloning or plasmid analysis are indicated by symbols defined at the bottom of the figure. The λKH100 DNA IS*5* left terminal *Taq*I fragment (λ:38,000–IS*5*: 335) has been cloned into the *Cla*I site of pHL200 in both orientations which correspond to the outside-in (t_{5o}) and large gene terminator (t_{5l}) respectively. Two sets of clones with different tetracycline resistance levels have been obtained: pHK1 and pHK2. A λKH100 DNA IS*5* right terminus containing *Taq*I–*Hin*dIII fragment (IS*5*:1,005–λ:38,291) corresponding to the small gene terminator, t_{5s}, has been cloned into the *Cla*I–*Hin*dIII gap of pHL200. The resulting hybrid plasmid pHK3 showed a strong reduction of the tetracycline resistance level. The control clones have been constructed in a similar way using two φ80 DNA *Taq*I fragments for insertion into the *Cla*I site of pHL200:φ80:40,107–40,514 which contains no transcription signal, but two long φ80 − *O*, *P* reading frames either in sense or in antisense orientation (pHK4, pHK5), which extend into and probably interfere with the tetracycline resistance initiation region, and φ80:39,406–39,584 containing the t_0 signal in direct or inverse orientation (pHK6, pHK7). The position of the IS*5* transcriptional terminator signals and the adjacent λ sequences are indicated according to Fig. 1.

Note added in proof: Since submission of this paper two independent sequence determinations of IS*5* DNA from different isolates have been published. They gave an identical[36] or nearly identical[37] (two deviations) result.

Received 16 March 1981; accepted 18 March 1982.

1. Starlinger, P. *Plasmid* **3**, 241–259 (1980).
2. Calos, M. P. & Miller, J. H. *Cell* **20**, 579–595 (1980).
3. Kopecko, D. J. *Progr. molec. subcell. Biol.* **7**, 135–234 (1980).
4. Reif, H. J. & Saedler, H. *Molec. gen. Genet.* **137**, 17–28 (1975).
5. Kleckner, N., Reichardt, K. & Botstein, D. *J. molec. Biol.* **127**, 89–115 (1979).
6. Davidson, N., Deonier, R. C., Hu, S. & Ohtsubo, E. in *Microbiology 1974* (ed. Schlessinger, D.) 56–68 (American Society for Microbiology, Washington DC, 1975).
7. MacHattie, L. A. & Jackowski, J. B. in *DNA Insertion Elements, Plasmids, and Episomes* (eds Bukhari, A. I., Shapiro, J. A. & Adhya, S. L.) 219–228 (Cold Spring Harbor Laboratory, New York, 1977).
8. Kleckner, N. & Ross, D. G. *Cold Spring Harb. Symp. quant. Biol.* **43**, 1233–1246 (1979).
9. Iida, S., Meyer, J. & Arber, W. *Cold Spring Harb. Symp. quant. Biol.* **45**, 27–43 (1981).
10. Jordan, E., Saedler, H. & Starlinger, P. *Molec. gen. Genet.* **132**, 353–363 (1968).
11. Adhya, S. L. & Gottesmann, M. E. *A. Rev. Biochem.* **47**, 967–996 (1978).
12. Lusky, M., Kröger, M. & Hobom, G. *Cold Spring Harb. Symp. quant. Biol.* **45**, 173–176 (1981).
13. Blattner, F. R. *et al. Virology* **62**, 458–471 (1974).
14. Szybalski, W. in *DNA Insertion Elements, Plasmids, and Episomes* (eds Bukhari, A. I., Shapiro, J. A. & Adhya, S. L.) 583–590 (Cold Spring Harbor Laboratory, New York, 1977).
15. Chow, L. T. & Broker, T. R. *J. Bact.* **133**, 1427–1436 (1978).
16. Lieb, M. *Gene* **12**, 277–280 (1980).
17. Glansdorff, N. & Sand, G. *Biochim. biophys. Acta* **108**, 308–311 (1965).
18. Crabeel, M., Charlier, D., Cunin, R. & Glansdorff, N. *Gene* **5**, 207–232 (1979).
19. Maxam, A. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564 (1977).
20. Sauer, R. T. *Nature* **276**, 301–302 (1978).
21. Landsmann, J., Kröger M. & Hobom, G. *Gene* (in the press).
22. Kahmann, R. & Kamp, D. *Nature* **280**, 247–250 (1979).
23. Dhar, R., McClements, W. L., Enquist, L. W. & van de Woude, G. F. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3937–3941 (1980).
24. Shoemaker, C. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 3982–3936 (1980).
25. Shine, J. & Dalgarno, L. *Proc. natn. Acad. Sci. U.S.A.* **71**, 1342–1346 (1974).
26. Rosenberg, M. & Court, D. *A. Rev. Genet.* **13**, 319–353 (1979).
27. Rak, B., Lusky, M. & Hable, M. *Nature* **297**, 124–128 (1982).
28. Ghosal, D., Sommer, H. & Saedler, H. *Nucleic Acids Res.* **6**, 1111–1122 (1979).
29. Klaer, R. *et al. Cold Spring Harb. Symp. quant. Biol.* **45**, 215–224 (1981).
30. Grindley, N. & Joyce, C. M. *Cold Spring Harb. Symp. quant. Biol.* **45**, 125–133 (1981).
31. Ohtsubo, H. & Ohtsubo, E. *Proc. natn. Acad. Sci. U.S.A.* **75**, 615–619 (1978).
32. Johnsrud, L. *Molec. gen. Genet.* **169**, 213–218 (1979).
33. Hobom, G. & Lusky, M. in *Mechanistic Studies of DNA Replication and Genetic Recombination* (ed. Alberts, B.) 231–255 (Academic, New York, 1980).
34. Langridge, J., Langridge, P. & Bergquist, P. L. *Analyt. Biochem.* **103**, 264–271 (1980).
35. Hobom, G. *et al. Cold Spring Harb. Symp. quant. Biol.* **43**, 165–178 (1979).
36. Schoner, B. & Kahn, M. *Gene* **14**, 165–174 (1981).
37. Eugler, J. A. & van Bree, M. P. *Gene* **14**, 155–163 (1981).

Molecular structure of a new family of ribonucleases

Y. Mauguen*, Robert W. Hartley†, E. J. Dodson‡, G. G. Dodson‡, G. Bricogne§||, Cyrus Chothia§ & A. Jack§¶

* Centre Pharmaceutique et LURE, Université Paris-Sud, 92290 Chatenay-Malabry, France

† Laboratory of Nutrition and Endocrinology, National Institute of Arthritis, Diabetes and Digestive and Kidney Diseases, Bethesda, Maryland 20205, USA

‡ Department of Chemsistry, University of York, Heslington, York YO1 5DD, UK

§MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Bacillus amyloliquefaciens produces a ribonuclease[1] (barnase), the function of which is probably the digestion of external RNA: it is excreted by the bacillus and within the cell its action is inhibited by a protein of ~89 residues to which it binds with high affinity[2]. Determination[3] of the amino acid sequence of the ribonuclease, which is a monomer of molecular weight 12,382 consisting of 110 residues, has revealed its homology with other prokaryotic and eukaryotic ribonucleases[4–6]. We have now determined the atomic structure of barnase by X-ray crystallographic analysis. We report that its structure includes a large central β-pleated sheet and two α-helices. The arrangement of these secondary structures is different from that found in bovine pancreatic ribonuclease.

Crystals of barnase have the space group $P3_2$ and the hexagonal unit cell dimensions: $a = b = 59.0$ Å, $c = 81.6$ Å. There are three molecules in the asymmetric unit. X-ray diffrac-

|| Present address: Department of Biochemistry, Columbia University College of Physicians and Surgeons, 630 West 168 Street, New York, New York 10032, USA.

¶ Deceased.

0028–0836/82/190162—03$01.00

Table 1 X-ray data and heavy atom refinement for barnase

Crystal	Data agreement F (R_M)	Heavy atom parameter agreement R_1	R_2	R_3	Heavy atom reaction sites (no. in asymmetric unit)	Heavy atom contributions and errors	$4\sin^2\theta/\lambda^2$ 0.00	0.04	0.08	0,12	0.16
Native	0.04										
Gold I	0.03	0.43	0.18	0.09*	His 102 (3)	⟨FH⟩	262	207	169	136	110
						⟨EH⟩	62	62	79	66	50
Gold II	0.04	0.44	0.19	0.08*		⟨FH⟩	—	—	—	148	122
						⟨EH⟩	—	—	—	77	65
Iodide	0.05		0.15	0.12†	Tyr 13 (6)	⟨FH⟩	—	134	100	72	54
					Tyr 17 (2)	⟨EH⟩	—	134	126	98	80

$$R_1 = \frac{\sum||FHLE| - |FH_{calc}||}{\sum|FHLE|}; \quad R_2 = \frac{\sum||FPH| - |FP||}{\sum|FPH|}; \quad R_3 = \frac{\sum||FPH| - |FPH_{calc}||}{\sum|FPH|}$$

where $FHLE$ is the experimental value of the heavy atom contribution, FH_{calc} is the calculated heavy atom contribution[7], $|FP|$ is the observed structure amptitude of the protein, $|FPH|$ the observed structure amptitude of the derivative, and $|FPH_{calc}|$ the derivative structure amptitude constructed by combining the calculated heavy atom contribution and the protein structure factor. $\langle FH \rangle$, r.m.s. value of FH_{calc} in the appropriate range of $4\sin^2\theta/\lambda^2$; $\langle EH \rangle$, r.m.s. lack of closure.

* $|FPH_{calc}|$ was constructed using the symmetry-averaged phases.

† $|FPH_{calc}|$ was constructed using the isomorphous phases defined by the gold derivative.

tion data were collected to 2.5 Å spacing on a Hilger and Watts automatic 4-circle diffractometer using the native enzyme and one derivative, which was prepared by soaking the crystals in gold cyanide. This derivative crystal diffracted well and the anomalous scattering effects produced by the gold were recorded. Owing to differences in the gold substitution, the data for the two crystals used in the case of the gold derivative were treated separately throughout. A second heavy atom derivative was prepared by iodination of the crystals and a 2.2 Å data set collected photographically at LURE using synchrotron radiation selected at $\lambda = 1.40$ Å. The anomalous scattering in this data set was not measured.

Three gold atom positions were identified from the difference Patterson function with coefficients $(|F_P| - |F_{PAu}|)^2$ and their parameters were refined by least-squares minimization to the experimental heavy atom contribution (F_{HLE})[7]. Iodine positions were determined from the difference Fourier map (coefficients $||F_P| - |F_{PI}||$) phased by the gold derivative[7]. A phase set was calculated from the two derivatives using Blow and Crick's combination procedure[8]. Details of the heavy atom refinement and phase calculations are given in Table 1.

We were unable to interpret the electron density calculated with the combined isomorphous phases although there were regions in which the protein's main-chain and side-chain structure could be traced in detail. There was also an encouraging similarity in the electron density of each of the three independent molecules. The non-crystallographic symmetry relations between the molecules were first determined from the gold positions and then refined by correlating the electron densities within spheres centred approximately on each molecule, using

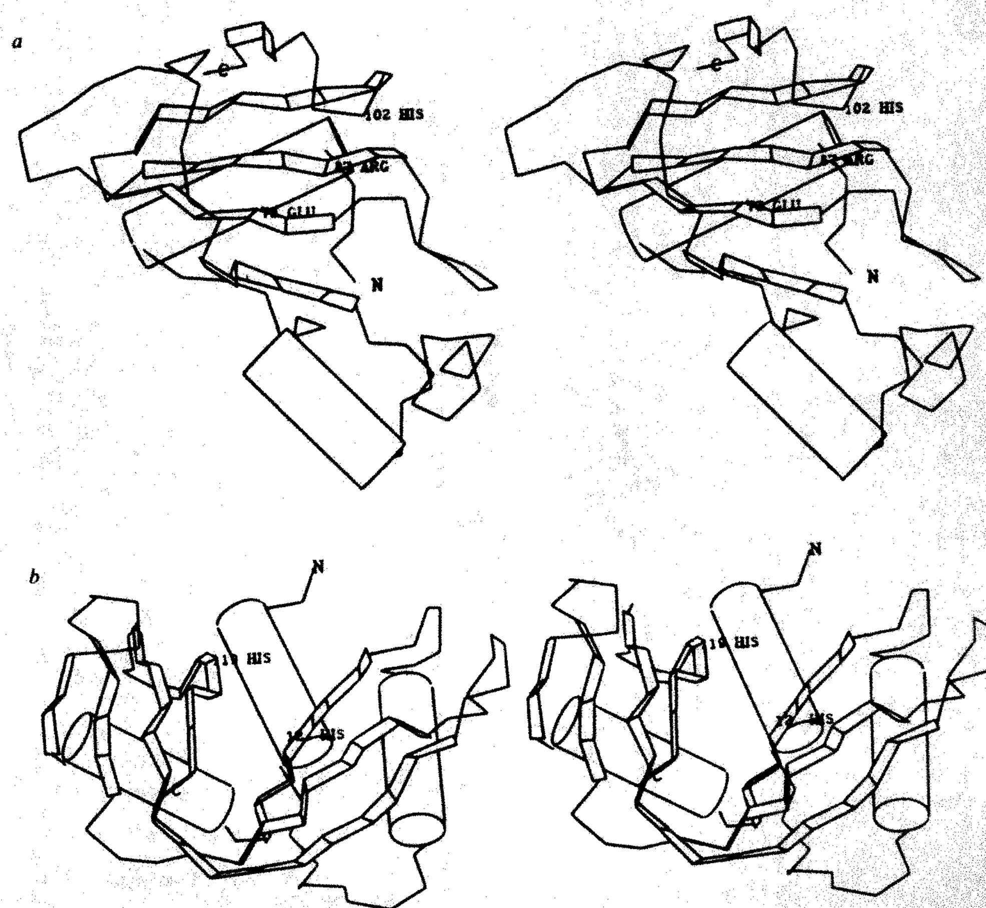


Fig. 1 Stereo pairs of schematic drawings of *a*, barnase and *b*, bovine pancreatic ribonuclease A. Each vertex represents the position of a $C\alpha$ atom; α-helices are shown as cylinders and the strands of the β-sheet as ribbons. The connecting loops are represented by lines. The active site of barnase includes Glu 73, Arg 87 and His 102 and is shown in more detail in Fig. 3. The active site of bovine pancreatic ribonuclease A includes His 12 and His 119. This figure was drawn from the atomic coordinates using the computer program of Lesk and Hardman[20].

a least-squares superposition method[9]. The mean of the correlation coefficients between the different molecules was 0.60.

The molecular envelope was defined from the isomorphous map, in which the electron density was averaged using the refined non-crystallographic symmetry operations. The map was sectioned in two perpendicular directions to ensure accurate tracing of the envelope. This was important to avoid the need for recalculating the envelope during the refinement calculations.

The electron density map produced with the symmetry-averaged phases[9,10] was readily interpreted. All but two residues have sensible density, which in some cases is very well defined, and their coordinates are now being prepared for refinement calculations.

The polypeptide chain of barnase is shown in Fig. 1*a*. The secondary structure of the enzyme consists of a central, five-stranded, twisted antiparallel β-sheet (residues 50–55, 70–75, 85–91, 94–101 and 106–108) and two α-helices (residues 6–18 and 26–34). The first helix packs against the face of the β-sheet in the normal manner[11] and the second helix packs against its edge. Bovine pancreatic ribonuclease also consists of a large central β-sheet with adjacent α-helices[12,13] but it differs from barnase in that its β-sheet is not only twisted but also bent in the middle, giving it a V shape (Fig. 1*b*). In addition, the residues that form the active site in the two molecules originate in parts of the structures that are not analogous (see below) and give different base specificities: barnase specific for purines[14] and pancreatic ribonuclease for pyrimidines[12].

Extracellular ribonucleases from *Bacillus intermedius*, *Streptomyces erythraeus* and from the fungi *Aspergillus oryzae* and *Ustilago sphaerogena*, have amino acid sequences that are homologous to that of barnase[4–6]. Figure 2 shows that residues 30–110 of barnase are between 82 and 20% homologous with the four other sequences. There are five positions at which the residues are invariant; in barnase these positions are occupied by Leu 42, Glu 73, Ser 80, Arg 87 and His 102 (Fig. 2). The residues at three of these positions have been shown to be involved in the ribonuclease activity of the proteins: the His in barnase and ribonucleases St and T_1, the Arg in ribonucleases St and T_1, and the Glu in ribonuclease T_1 (refs 6, 15, 16).

```
     30        40        50
Ba   AQALGWVASKGNLADVA-P-GKSIGGDIFSNR
Bi   ASALGWVASKGDLAEVA-P-GKSIGGDVFSNR
St   ATDTGY-----ELIEKGGPYPYPEDGTVFENR
T1   AQAAGY-----QLHEDGETVGSNSYPHKYNNY
U2   -----------GALDDVARPDG-DNYPHQYYDE
     --α--                   ---β---

     60        70   *          80
Ba   EGKLPGKSG--RTWREADINYT----SGF-----
Bi   EGRLPSAGS--RTWREADINYV----SGF-----
St   EGILPDCAE--GYYHEYTVKTP----SGDD----
T1   EGFDFSVSS---PYYEWPILSSGDVYSGPG----
U2   ASDQITLCCGPGSWSEFPLVYNGPYYSSRDNYVS
                ---β---

         *  90        100 *      110
Ba   RNSDRILYSSD-WLIYKTTDHYQT-FTKIR----
Bi   RNADRLVYSSD-WLIYKTTDHYAT-FTRIR----
St   RGARRFVVGDG-GEYFYTEDHYES-FRLTIVN--
T1   SGADRVVFNEN-NQLAGVITHTGA-SGNNFVECT
U2   PGPDRVIYQTNTGEFCATVTHTGAASYDGFTECS
        ---β---     ----β----   -β-
```

Fig. 2 An alignment of the amino acid sequences of the extracellular ribonucleases from *B. amyloliquefaciens*[3] (Ba); *B. intermedius*[5] (Bi); *S. erythraeus*[6] (St); *A. oryzae*[21] (T_1); and *U. sphaerogena*[22] (U_2). The numbering used here is that of barnase so that the positions of mutations, insertions and deletions can be seen in Fig. 1*a*, 3. Residues that are identical in two or more sequences are boxed and homologous residues that are part of the active site (see text) are indicated by asterisks. The regions that form secondary structures in barnase are indicated by 'α' for an α-helix and by 'β' for the strand in the β-sheet. The alignment shown here has been revised slightly from that given in ref. 4 in the view of the barnase structure. The residues that form the first part of the five sequences show little overall homology and large deletions. It has been shown that the alignment of such sequences derived by maximizing the match of the few similar or identical residues can give results that are structurally incorrect[23].

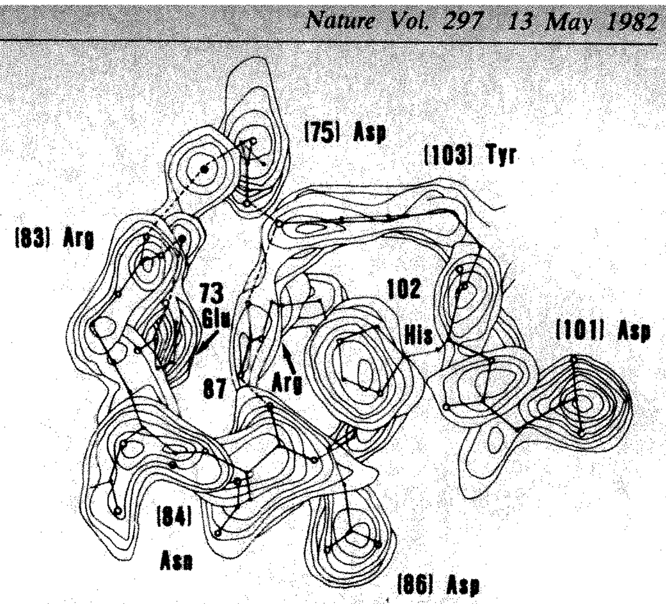


Fig. 3 A composite view of the electron densities at the active site, with the protein atomic coordinates superimposed. Two water molecules are indicated by solid circles.

Inspection of the barnase structure shows that residues Glu 73, Arg 87 and His 102 cluster in one region of the molecules, therefore this region must form the active site (see Figs 1*a*, 3).

The catalytic role of hisitidine 102 has been confirmed by the finding that gold cyanide, which reacts with imidazole N_D in the crystal, suppresses the enzyme's activity. Removal of the gold cyanide by dialysis restores fully the catalytic activity of the enzyme (R.W.H., unpublished results).

In barnase, the active-site residues are in a shallow depression and come solely from the C-terminal half of the molecule. Their chemical and structural organization differs from that at the active site of bovine pancreatic ribonuclease A, where the principal catalytic residues are two histidines, residues 12 and 119, which lie in a deep cleft (Fig. 1*a*, *b*).

Work is in progress to elucidate the catalytic mechanism of barnase and to provide a structural interpretation of previous experiments on its folding[16]. X-ray analyses of the ribonucleases T_1 and St are in progress elsewhere[17–19] and comparison of their structures with that of barnase should allow an understanding of how this family of ribonucleases evolved.

We acknowledge with gratitude the use of the synchrotron radiation facilities at LURE and the help given by Peter Moody, Federico Giordano and Christopher Hill during the analysis.

Note added in proof: Since submitting this paper we have learnt of the structure determination at 2.5 Å resolution of ribonuclease St (Yamamoto, Y. *et al.* 9th Symposium on Nucleic Acid Chemistry, Tokyo, October 1981) and of ribonuclease T (Heinemann, U. and Saenger, W., personal communication).

Received 18 January; accepted 19 March 1982.

1. Nishimura, S. & Nomura, M. *Biochim. biophys. Acta* **30**, 430–433 (1958).
2. Hartley, R. W. & Smeaton, J. R. *J. biol. Chem.* **248**, 5624–5626 (1973).
3. Hartley, R. W. & Barker, E. A. *Nature new Biol.* **235**, 15–16 (1972).
4. Hartley, R. W. *J. Molec. Evol.* **15**, 355–358 (1980).
5. Aphanasenko, G. A., Dudkin, S. M., Kaminiv, L. B., Leshchiukaya, I. B., Leshchiukaya, I. B. & Severin, E. S. *FEBS Lett.* **97**, 77–80 (1980).
6. Yoshida, N., Sasaki, A., Rashid, M. A. & Otsuka, M. *FEBS Lett.* **64**, 122–125 (1976).
7. Dodson, E. J. *Crystallographic Computing Techniques*, 260–268 (ed. Ahmed, F. R.) (Munskgaard, Copenhagen, 1975).
8. Blow, D. M. & Crick, F. H. *Acta crystallogr.* **12**, 794–802 (1959).
9. Bricogne, G. *Acta crystallogr.* **A32**, 832–847 (1976).
10. Bricogne, G. *Acta crystallogr.* **A30**, 395–401 (1974).
11. Janin, J. & Chothia, C. *J. molec. Biol.* **143**, 95–128 (1980).
12. Richards, F. M. & Wyckoff, H. W. *Enzymes*, 647–806 (1971).
13. Carlisle, H. C., Palmer, R. A., Mazmudar, K. S., Gorinsky, B. A. & Yeates, D. G. R. *J. molec. Biol.* **85**, 1–18 (1974).
14. Rushizky, G. W., Greco, A. E., Hartley, R. W. & Sober, H. *Biochemistry* **2**, 787–793 (1963).
15. Takahashi, K *J. Biochem., Tokyo* **80**, 1267–1275 (1976).
16. Hartley, R. W. *J. Biol. Chem.* **252**, 3252–3254 (1977).
17. Martin, P. D., Tulinsky, A. & Walz, F. G. Jr *J. molec. Biol.* **136**, 95–97 (1980).
18. Heinemann, U. *et al. Eur. J. Biochem.* **109**, 109–114 (1980).
19. Yamamoto, Y. *et al. J. molec. Biol.* **145**, 285–287 (1981).
20. Lesk, A. M. & Hardman, K. *Science* (in the press).
21. Takahashi, K. *J. biol. chem.* **240**, 4117–4119 (1965).
22. Sato, S. & Uchida, T. *Biochemistry* **145**, 353–360 (1975).
23. Delbaere, L. T. J., Brayer, G. D. & James, M. N. G. *Nature* **279**, 165–168 (1979).

BOOK REVIEWS

From Kharkov to Tel Aviv

John M. Charap

Refusenik: Trapped in the Soviet Union. By Mark Ya Azbel. Pp.513. UK ISBN 0-241-10633-8; US ISBN 0-395-30226-9. (Hamish Hamilton/Houghton Mifflin: 1982.) £9.95, $17.95.

IF YOU look up *refusenik* in a Russian dictionary, you will not find it there. After all, it is not an official Russian word, but has been improvised to describe someone who has applied to emigrate from the USSR to Israel, and whose application for a visa has been refused. What the word does not convey is the anguish and the courage of those who wait, often for years, trapped in a situation in which as individuals they face the power of a state at best indifferent, frequently hostile.

Mark Yakovlevitch Azbel is probably best known to many readers of this journal for his work on cyclotron resonance in metals. He learnt his physics at "the cradle of Soviet physics", Kharkov State University, where Lev Landau had established a world-famous research institute. From Kharkov he went to Chernogolovka, to the Landau Institute for Theoretical Physics, where he continued to rise to a position of prominence and international repute. Yet in December 1972 he took the fateful step of applying for a visa to emigrate to Israel, and became a *refusenik*. It was not until July 1977 that he was finally permitted to leave the USSR.

This remarkable book tells how it was that, notwithstanding his success and prospects for continuing advancement as a Soviet physicist, he became convinced that he could not remain in the Soviet Union. And it tells of the hazards and torments which dogged his steps from Kharkov to Tel Aviv. Azbel has written a moving and sincere autobiography. Any such work from the Soviet Union must be of interest, because of the light it sheds, sometimes no less illuminating for being oblique, on life in that country. For readers of *Nature* it will be of particular interest because of the view it gives of Soviet science, and indeed of some individual Soviet scientists. But to my mind the real importance of this book lies in its exposition, from the stand-point of one who experienced and suffered from it, of the corrupting pervasion of anti-Semitism in the USSR.

Azbel's parents were both physicians, who graduated in Kharkov in the terrible days of 1931 when famine stalked the land, the man-made famine which followed forced collectivization, and Mark was born the following year. As a four-year-old, left unsupervised whilst his parents worked six days a week, Mark first found out that he was different, that he was a Jew. Lenin had tried to eliminate the anti-Semitism which was so deeply rooted in the USSR — nowhere more than in the Ukraine. Stalin, as first Commissar for Nationalities, had made it in a certain sense impossible to be a Russian Jew, or a Ukrainian Jew. To be a Jew in the Soviet Union is not a matter of choice or conviction; it is rather a question of nationality, and if it says "Jew" on an internal passport, it cannot say "Russian" or "Ukrainian". From his first day at nursery school when his teacher greeted him with "Another zhid! Where do they all come from?" to his last few hours in Moscow when the cold insulting indifference of the officials at the airport made his departure a nightmarish farce, Azbel was time and again reminded of his difference, of his Jewishness.

Of course there were times when this awareness could recede into the background. But with dramatic force public affairs would make him, along with the whole of the Jewish community, aware of the precarious — even dangerous — status of the Jews. As a student he felt the tension which prevailed throughout the country during Stalin's last years.

> The Lysenko catastrophe had repercussions all over the Soviet Union affecting biologists in even the smallest, most remote institutions, obscure little places of minor importance. Now the laws against Cosmopolitanism were threatening artists, composers, writers. Russians were not wholly exempt from these attacks; but nine out of ten of the criminals were Jews.

Confronted by the menace and threat of disappearance, arrest by the KGB or worse, Azbel asked himself: how could one survive in this atmosphere? The question was not about physical survival, but moral survival. There were several possibilities: concentration on personal life; devotion to a struggle with the system; escape into pure science. But

> I could not make any of these choices. I felt that my own way of life was to be in science, but as to friends: it seemed to me that it would be crippling, demeaning, to choose my friends on the basis of which friendships would be safe and which would not. To confine my loyalties to science, and to my fellow-scientists, for reasons of security; to cultivate indifference to everything and everyone else — I knew I couldn't do that. In a word, I simply made up my mind at that time that I would try to be a decent human being.

Stalin died, and the terrible threat implicit in the "Doctors' plot" receded. Azbel graduated, brilliantly, but not without politically motivated machinations against him which it required the intervention of Ilya Lifschitz to deflect — and not without his first encounter with the KGB, perhaps an attempt to obtain a hold on him for future exploitation. The heavy atmosphere lightened with Khruschev's short-lived thaw. Azbel married and moved to Moscow.

Also now in Moscow was Alexander Voronel, a close friend from student days in Kharkov and now a director of a laboratory in Dubna: they had much in common, and their mutual friendship and support for Andrei Sinyavsky, and especially Yuli Daniel, brought them both into the hands of the KGB when these two authors were put on trial in 1965. This brush with the authorities was more significant than the earlier one, and Azbel was lucky to avoid losing his post at the university.

The shifting attitudes within the Soviet Union during the late 1960s, responsive to events such as the Six-Day War and the Czech uprising, meant that when in 1970 news first began to filter through that some Soviet Jews had publicly expressed a wish to emigrate to Israel, Azbel was amongst those who began to recognize that to leave the Soviet Union was a real possibility. His

Contemporary Jewish Library

The start of active protest: four of the seven scientists who took part in the 15-day hunger strike, June 1973, in reaction to the refusal of the Soviet authorities to allow them exit visas. From left to right: Dan Roginsky, Alexander Voronel, Mark Azbel and Victor Brailovsky.

resolve was stiffened by the trial in Leningrad of Dymshitz and Kuznetsov and the demonstrations in the Visa Department offices in Moscow in 1971.

And so it was that in December 1972 he went to the Visa Department. He was not the first distinguished scientist to do so: that was probably Alexander Lerner who had applied a year earlier. Lerner is still a *refusenik* more than ten years later. Nor was he the last. Recent information suggests there are about 100 *refuseniks* with a scientific doctorate (roughly equivalent to DSc) in Moscow alone, another 200 with the degree of Candidate (roughly PhD). Voronel applied shortly before Azbel, and was allowed to leave at the end of 1974.

It must be understood that to apply to emigrate from the USSR is an act of the highest significance. It leads almost inevitably to social ostracism. And in most cases, certainly for scientists, it leads to dismissal from one's job. For a scientist this in turn means more than loss of livelihood (which can in turn lead to prosecution for "parasitism"); it means deprivation from contact with colleagues, from access to libraries and laboratories. It means "exclusion from science". To counter that exclusion, to provide mutual support and scientific sustenance, Voronel organized a seminar every Sunday noon at his apartment. These Sunday seminars continue to this day, despite a succession of attempts to disrupt or prevent them. They have multiplied: similar meetings are held in other Soviet cities, and in disciplines other than physics. The participants have received Western visitors in this way (a recent session learnt about Feigenbaum's work on frequency doubling and the transition to chaos): indeed they have held impressive international meetings. All of this has been under the watchful eyes of the KGB, and has been scrupulously within the law.

The legality (and to Western readers perhaps the normality) of private gatherings of scientists to discuss science has not prevented harrassment, and official attempts to disrupt their proceedings. These reached a climax when Voronel, Azbel and Victor Brailovsky decided in 1974 to hold an international scientific symposium, and again three years later when another international meeting, on collective phenomena in physics, was held to celebrate the fifth anniversary of the seminars. Both meetings were attended by numerous distinguished visitors from abroad. Both were heralded by oppressive KGB surveillance, questionings and indeed arrests — Azbel was himself imprisoned in 1974. After Azbel's emigration, Victor Brailovsky took over as chairman of the Sunday seminars from Azbel, who had succeeded Voronel. Brailovsky is now living in exile in Beineu, Kazakhstan.

Azbel's book speaks eloquently of the courage and resourcefulness of those scientists who persist in their legitimate aspirations to emigrate and of their motivation. It answers some but not all of the questions which must be raised about the reasons behind the Soviet policies towards them. It underlines their vital need for support from the international scientific community, which I believe to be an obligation upon us as scientists. The pursuit of science gives privileges and brings responsibilities: and it is not possible to separate those responsibilities from the wider relations between the pursuit of science and human affairs. This book should be read by all those who profess an interest in international scientific activities, and I would hope that would mean all scientists. □

John Charap is Head of the Physics Department at Queen Mary College, University of London. He attended the Moscow Sunday seminar in 1978, and is Chairman of the European Physical Society Advisory Committee on Scientific Freedom.

The quantum theory of space and time

C.J. Isham

Quantum Fields in Curved Space. By N.D. Birrell and P.C.W. Davies. Pp.340. ISBN 0-521-23385-2. (Cambridge University Press: 1982.) £27.50, $49.50.

THE problem of quantizing the gravitational field commands the attention of an increasing number of theoretical physicists who are attracted by the importance of the subject to studies of the early Universe, gravitational collapse, grand unification of all fundamental forces, and the Planck length structure of space and time. In spite of much effort a successful resolution remains elusive, and it seems natural to tackle first the simpler problem of the quantization of a field propagating in a curved, but fixed, background spacetime. It had been known for many years that a time dependent metric would transmute part of its energy through the medium of quantum particle production, but the subject became of widespread interest only after Hawking's announcement in 1974 of the production of particle pairs by the event horizon of a black hole.

The subsequent effort devoted to quantum field theory in a curved space has spawned a considerable literature of research papers, conference proceedings and review articles, but the text of Birrell and Davies is the first attempt to present the subject in a more permanent form. I can think of at least three quite different types of book that might be published under this title with the contents reflecting the individual approaches that have been developed by different researchers. The present authors have made a determined and commendable effort to cite most of the important papers that have appeared, but the final treatment sharply reflects their own personal contributions and methodology. Quantum field theory is introduced with heuristic methods involving mode function expansions; there is no discussion of the mathematical problems associated with such sums or of their resolution via the operator algebra techniques that form the backbone of most rigorous investigations into quantum fields with external sources. This simpler approach does however have the asset of rendering the book accessible to a postgraduate student who has attended any standard first course on quantum field theory, and the assumed knowledge of general relativity is no more demanding.

A significant application of the subject is to the physics of the early Universe; for example it has been suggested frequently that an initial anisotropy could be removed by the quantum production of particle pairs. Birrell and Davies are especially interested in this problem and provide a comprehensive discussion of quantum field theory in various cosmological backgrounds. Of particular value here is the careful account of the resolution, using model particle detectors, of the conceptual problems that arise concerning the operational status of a quantum particle.

Almost all work in quantum field theory is plagued by the infamous ultraviolet divergences and the addition of a curved background tends, if anything, to make matters worse. A particularly disturbing feature is the appearance of infinities in the matrix elements of the energy momentum tensor which, in a semiclassical limit, forms the source of the gravitational field. The regularization and renormalization of these divergences is a subtle problem and is discussed at length with most of the established techniques being given a respectable hearing; applications include an authoritative account of the Hawking effect and of the ensuing use of the stress tensor to isolate the origin of the particle pairs.

Professor Davies has rightfully earned a high reputation as an author of informative and well written semi-popular science books and this research monograph is presented in a similarly clear and attractive style. Quantum gravity is an active subject and it would be difficult to predict that all sections of the text will seem equally relevant in ten years time; nevertheless, the authors are to be congratulated on producing a timely work that should help to stimulate interest in this fascinating branch of theoretical physics. □

C.J. Isham is Reader in Theoretical Physics at Imperial College, University of London.

Mycology: a successor to Cochrane

John D. Weete

Fungal Physiology. By David H. Griffin. Pp.383. ISBN 0-471-05748-7. (Wiley: 1981.) £25.45, $43.25.

DURING the past few years several excellent, advanced-level monographs have appeared on aspects of the biochemistry and physiology of fungi, reflecting some 20 years of continual advance in our understanding of these subjects. Yet there has been no successor to Cochrane's general introductory text, *Physiology of Fungi,* published by John Wiley in 1958, and long out of date. Professor Griffin has now provided such a text with *Fungal Physiology.* His coverage emphasizes those physiological processes typically associated with fungi, and stresses the experimental rather than historical aspects of the subject.

The opening three chapters, reviewing the fungal thallus, and the chemical composition and structural organization of fungus cells, are followed by an outline of the major processes occurring within fungal cells. On balance I believe more emphasis could have been given in this latter chapter to metabolism and the treatment strengthened by greater coverage of the relationships between metabolic change, growth and reproduction. The author then turns to various aspects of vegetative growth, and in the next three chapters considers the uptake of nutrients, the chemical requirements for growth and the relation of the physical environment to growth. Chapters 9 to 12 deal with sporogenesis (environmental factors, biochemistry), dormancy and germination, and the role of hormones in sexual reproduction. Here, Professor Griffin has chosen *Dictyostelium discoideum* rather than a yeast or filamentous fungus to illustrate the biochemistry of spore formation. He provides ample evidence indicating that fungi of different classes seem to have developed different hormonal systems for regulating sexual reproduction. The chapter on syngamy (Chapter 12) is organized on the basis of fungal groups and the specific hormones associated with each group are discussed.

Understanding specific biochemical processes can be aided by the use of inhibitors; in this context, inclusion of a chapter devoted to fungicides is noteworthy because of the significance of many of these compounds in the agrochemical and biomedical fields as chemical agents for controlling pathogenic fungi. The final chapter, "Fungal Attack Mechanisms", is concerned with fungal toxins, exoenzymes, fungal responses to environmental stimuli and fungal interactions (non-pathogenic) with other living systems. Fungal genetics *per se* is not covered, but genetic studies are discussed where they are relevant to other topics, particularly fungal symbiotic relationships. The importance of using mutants in mycological research is implied by the citation of examples throughout the book.

The book is well organized; each chapter begins with an introductory statement which puts the subject under consideration into perspective, ends with a brief summary, and is appropriately referenced. Topic discussions are well illustrated with light and electron micrographs, tables and, where pertinent, graphs and chemical structures. *Fungal Physiology* is suitable as a text for a general course serving advanced undergraduate and graduate students, and also provides a handy reference work for researchers seeking general background information. The appearance of this book is both significant and timely; I highly recommend it. □

John Weete is a Professor in the Department of Botany, Plant Pathology and Microbiology at Auburn University, Alabama.

View of laser fusion

David W. Forslund

Physics of Laser Driven Plasmas. By Heinrich Hora. Pp.317. ISBN 0-471-078880-8. (Wiley: 1981.) £28.95, $49.15.

THE field of laser fusion and the physics of laser driven plasmas has developed rapidly in the past decade and it has proven difficult for research workers to keep up with progress. Particularly lacking have been good reviews and good textbooks which would help introduce new people to this extremely complex subject. In this recent attempt to meet such a need, Professor Hora has sought to derive the basic physics of the interaction of intense laser light with matter from first principles. He attempts to lead the reader through the derivation of the fundamental plasma physics, kinetic theory, hydrodynamics and an expression for the nonlinear force arising from intense laser light, and follows this with discussion of the numerous applications of the nonlinear force from its role in parametric instabilities to its use in the implosion of pellets.

The book suffers from a number of flaws, however, which prevent it from being useful either to the student or as a handbook for the research worker. First, the discussions of most of the phenomena are unduly complex, confused and full of errors, making the text extremely difficult to read. As a typical example, the derivation of the plasma frequency and Debye length on pages 22–25 introduces

the unclear and irrelevant ideas of "cell distance", "refractive index" and "electrostatic explosion of ions".

Throughout the book Professor Hora seems to be forcing his own conceptions on the reader rather than trying to properly survey the field. As a consequence the book gives a very unbalanced view. For example, only one paragraph is devoted to the production of high energy electrons, which continues to be such a major difficulty in laser fusion. Consistent with this neglect, the major problem of electron transport is totally ignored.

More than one-third of the book is spent on calculating the ponderomotive force and its effects on hydrodynamics, which leads, in the conclusion of Chapter 13, to the claim that all the problems of hot electrons in experiments can be removed by utilizing "the nonlinear force". As an alternative to the usual ablative compression approach which suffers from electron preheat, Professor Hora suggests

> . . . the very fast nonlinear force pusher will accelerate the plasma before the hot electrons can be generated. This has been confirmed by extensive numerical studies and can be the way out of the present difficulties of laser fusion.

Apparently, Professor Hora genuinely believes that a short pulse intense laser at the upper limit of already achieved intensities will not produce energetic electrons. It is remarkable that, although this simple target approach is at least five years old, no major experimental programme has sought to make use of it. Evidently, of the many proposals to reduce the preheat problem, Professor Hora's has been judged to be the least likely to succeed. □

David W. Forslund is a Fellow of the Los Alamos National Laboratory, New Mexico.

Seismologists look at arms control

Peter D. Marshall

Identification of Seismic Sources — Earthquake or Underground Explosion. Edited by Eystein S. Husebye and Svein Mykkeltveit. Pp.876. ISBN 90-277-1320-0. (Reidel: 1981.) Dfl. 195, $98.

In 1958 a Committee of Experts met in Geneva to study the possibility of verifying a treaty to ban the testing of nuclear devices. Under the treaty, eventually signed in 1963, the USA, USSR and UK stopped all but underground testing — although underground explosions could be detected by their seismic effects, uncertainty in discriminating between explosions and the many earthquakes of equivalent magnitude meant that a treaty to ban them was politically unacceptable. This uncertainty in discrimination has occupied forensic seismologists for over 20 years. Their latest views are aired in these proceedings of a NATO-sponsored Study Institute, convened to assess progress since systematic research began with the Vela Uniform programme in the USA and Tabor Pluto in the UK.

These programmes led to instrumental improvements and to the use of digital computers to increase the accuracy of detection and hypocentre location. Many more earthquakes could then be distinguished from explosions because of their greater depth; in addition, the spectral differences between earthquake and explosion signals led to identification by their body to surface-wave ratios. These improvements opened up the possibility of teleseismic rather than the short-range monitoring available in 1958, thereby allowing a potential reduction in the number of stations required to verify a treaty. However problems continued to arise: the increasing number of explosions in both seismic and aseismic regions brought differences in transmission paths to the attention of seismologists and provided much of the impetus for modelling seismograms. This in turn highlighted the loss of information suffered by the narrow recording bands in common use, and thought and effort was successfully brought to bear on digital recording and processing in much wider bands.

By 1976 many countries had initiated seismological discrimination studies and the UN Committee for Disarmament took advantage of this wider interest to set up a multinational Ad Hoc Group of Scientific Experts to make recommendations on how a Comprehensive Test Ban Treaty (CTBT) might be monitored. Soon afterwards, CTBT negotiations were re-opened between the USA, USSR and UK; these demonstrated a need to supplement teleseismic stations with regionally sited stations in order to achieve the verification levels required.

At this point, groups in the USA and Norway proposed a discussion of recent developments in discrimination studies under NATO sponsorship. The resulting proceedings of the Study Institute provide seismologists with a reference book of 45 papers, 20 by invited key lecturers.

The scene is set by Douglas's objective review which includes a comprehensive account of diagnostic criteria. Effective and stable diagnostic criteria depend on a knowledge of earthquake and explosive sources. Not surprisingly little is known about the earthquake source but strangely, as Rodean shows, much the same applies to nuclear explosions even after 20 years of underground testing. Any technique which defines their parameters thus becomes a significant diagnostic aid. One in which the seismic moment tensor is used to represent the source is discussed, among others, by Doornbos, clarifying what was becoming a rather confused topic and one turning out to be well worth further exploitation even though the method is bedevilled by transmission path effects. Waveform synthesis, discussed by both Harkrider and Kennett, may however one day lead to a solution of the problem of path effects. Of the seven papers on the influences of attenuation and scattering (both path effects) on the waveforms, that by Aki demonstrates the dramatic effects of even small velocity fluctuations within the lithosphere beneath a recording station.

In spite of the uncertainties, a seismologist must ultimately make a decision: earthquake or explosion? Much may depend on his decision, so the more independent observations available to him the better and Tjostheim reports on how multidimensional discrimination techniques may be applied to teleseismic data. The technique can also be applied to regional diagnostic data of the type described by Blandford. Finally, an indication is given of the kind of network which would need to be established with the appropriate instrumentation, automatic processing, data transmission facilities and data centres capable of promptly handling enormous quantities of data.

In 1958 Sir William (later Lord) Penney described seismology as a "stone-age science". The contents of this book demonstrate that this is no longer true, and the organizers of the meeting are to be congratulated on their initiative. Nonetheless problems remain. Seismic decoupling of explosions, explosion-like earthquakes, the detection of small explosions among earthquake signals and, above all, inadequate knowledge of the structure of the Earth all contribute to the difficulties of concluding a CTBT. Seismologists and physicists anxious not to rediscover the wheel will find in this volume the sources on which to build their new research programmes. □

Peter D. Marshall is a Principal Scientific Officer with the Ministry of Defence, Blacknest, Berkshire.

● A revised, paperback edition of *The Question of Animal Awareness* by Donald R. Griffin has been published by William Kaufmann Inc., price £6.20, and is available through W.H. Freeman. The original edition was reviewed in *Nature* **266**, 792; 1977.

● A.I. Sabra's *Theories of Light, from Descartes to Newton*, first published by the Oldbourne Book Co. in 1967, has been reissued by Cambridge University Press in both hardback and paperback. The book is essentially a reprint of the original edition, but some minor corrections and a new bibliography have been included. Price is hbk £20, pbk £6.95.

nature

20 May 1982

Good start on strategic arms

President Ronald Reagan's proposals for a new bilateral negotiation on strategic arms control are welcome. Will the Soviet Union listen? And will the Administration follow through?

President Reagan appeared to take a stride towards serious strategic arms control on 9 May when he made his first strategic arms control proposal to the Soviet Union and called for deep cuts in each side's arsenals. While the proposal deserves applause as a first step, it is like the proverbial horse that has been led to water but has not yet drunk. The Soviet Union has yet to respond. And unless it is followed by clear, realistic and consistent effort by the Administration in the coming months, even years, the progress Reagan has made will be undone.

The proposal, transmitted to the Soviet Union in detailed form at the same time as the President outlined it in a speech at his *alma mater*, Eureka College in Illinois, focuses on reducing the Soviet advantage in long-range intercontinental ballistic missiles (ICBMs) by asking that each side should reduce to 5,000 the number of its ballistic missile warheads, no more than half of which could be on land-based ICBMs. At the latest count, the Soviet Union may have 6,300 ballistic missile warheads, of which 5,100 are on land-based ICBMs. The United States has 7,500 ballistic missile warheads, of which only 2,154 are on land-based ICBMs. One effect of the proposal is that the Soviet Union would have to reduce its land-based ICBM warheads from 5,100 to 2,500, and so forgo all or most of its force of SS-18 giant "heavy" missiles, which could have up to 3,000 high-precision warheads and which could now, in theory, knock out most of the US force of land-based ICBMs in a single blow. The United States, which invests a smaller fraction of its strategic forces in land-based ICBMs, is at present below the 2,500 warhead ceiling in the President's proposal and could comply merely by retiring some of its older ICBM-carrying submarines without sacrificing much in force capability. So much for phase one.

The President also proposed a second phase of negotiations in which both sides would strive for equality by another measure, ballistic missile throw-weight, "at less than current American levels". Throw-weight is the weight of a missile's payload including warhead. This notion has grown out of an argument (backed by Eugene V. Rostow, director of the Arms Control and Disarmament Agency, and by Defense Secretary Caspar Weinberger) that an aggregate index of throw-weight would be a better basis for measuring current and future force levels than merely counting holes in the ground or launch tubes on submarines, as in Salt II. The President apparently resisted using this formula for the first phase of negotiations because the joint chiefs of staff felt that weight-counts are less verifiable than actual warheads. Rostow now says that the connection between phase one and phase two will depend on the Soviet Union's reaction and counter-suggestions, but that phase one alone would achieve substantial reductions in throw-weight on the Soviet side, the area where its forces have an advantage.

So far, the plan has only been sketched in public, and the Soviet side has just received it. More details may not be forthcoming for some time, but the plan has some obvious weaknesses and some curious strengths. One weakness is that the proposal does not include nuclear warheads carried by aircraft in its warhead ceiling of 5,000. Instead, the proposal is said to put a separate ceiling on bombers, limiting each side to between 350 and 400. On the US side, this would include the long-range older B-52 bombers and the new B-1 but also the intermediate range FB-111s based in Europe, armed with nuclear weapons and capable of striking at targets in the Soviet Union. The Soviet Union has long insisted that the FB-111s should be counted as strategic weapons, a view which the United States has rejected in the past. In return for including the FB-111s under the bomber ceiling, the United States would now insist that the Soviet Union includes the "Backfire" bomber under its bomber ceiling. The range and deployment of the Backfire bomber was a contentious issue in Salt II — and its exclusion from the Salt II ceilings was one reason why that agreement was never ratified. In this connection, domestically, the Reagan plan could give the Administration a respectable way to back out of building large numbers of MX missiles, which have been touted as — but not proved to be — a way to reduce the vulnerability of land-based ICBMs, although the obvious way for the United States to reduce its total number of ballistic missile warheads from 7,200 to 5,000 would be to scrap some older submarines, another could be to cancel part of the MX programme. The Administration is already embarrassed that its promise a year and a half ago to set the MX issue straight forthwith has not yet borne fruit.

The Reagan plan therefore deserves a fair wind. It is to be hoped that the Soviet Union will not reject it outright, as it rejected President Carter's proposal for deep cuts in missile forces in 1977. But advocates of a nuclear freeze in the United States (see *Nature* 29 April, p.790) and Europe should not delude themselves that the President has taken a step towards their version of arms control. A curious fact about the politics of arms control in the United States is that both the left and the right favour deep cuts in nuclear forces, but for different reasons. The left favours cuts (and the extreme left favours unilateral cuts) as part of the continuing process of negotiation. The right, in the United States, favours proposals for deep cuts on the grounds that they are a way of limiting Soviet striking power substantially — and that they also challenge Soviet protestations of good intentions on arms control. At the worst, it is always possible to say "See, we told you so. They don't mean what they say".

As it happens, it is not yet clear how sincere even the Reagan Administration is about arms control, no matter what the President said in Eureka. In their 18-month partnership, Secretary of State Alexander Haig and Secretary of Defense Caspar Weinberger have shown that they have radically different views of the US–Soviet military balance. As a result of these and other differences, often aired publicly, the Administration has yet to convince countries abroad (and even itself) where it stands on some key issues. Yet what is critical to the success of the strategic arms negotiations is not this opening flourish but whether, in following up, the Administration can speak with one voice and settle down to serious talks, taking Soviet rhetoric in its stride and keeping policy on a single course, rather than zigzagging and confusing the Soviet leadership. In short, it remains to be seen whether the horse will take a drink.

Unfortunately it will be several months before these issues can be clarified. Mr Brezhnev, having had the Soviet Union complain that President Reagan was "warmongering" by choosing to attend the NATO meeting planned for June and not the opening session of the Special Session on disarmament of the United Nations Assembly, has now ducked the United States offer (or challenge) of a meeting on that occasion. Instead, the two men may now meet on neutral ground sometime in October. The delay

0028-0836/82/200169-02$01.00

is probably to be welcomed. It will give the US Administration a chance to put its act together, and the Soviet Union a chance to resolve the delicate problem of continuity in strategic arms negotiations during the period when Mr Brezhnev will be succeeded by somebody else. And the delay will also be a chance to decide how the negotiations now proposed should be linked with those formally under way, but almost stalled, on the limitation of nuclear weapons in Europe. The United States has always intended that the two processes should be merged, as it is sensible that they should be. For only then will President Reagan's "zero option" seem secure. By October, the Soviet Union should be ready to soften the negotiating position it has repeated since last November — that it will withdraw some of its mobile missiles if the United States will agree not to introduce the Pershing II and cruise missiles.

French new world?

France wants next month's economic summit to talk technology. It should not expect too much.

The summer now begun in the Northern Hemisphere will be metaphorically long and hot. Several important issues fall to be decided, or at least to be given an airing. Strategic arms control (see above) is one. The South Atlantic (where it will be winter) is another. Next month, there will also be the Special Session of the United Nations on Disarmament in New York. Before then, President Reagan will have made the rounds of several West European capitals and will have sat in on two crucial if lesser summit-meetings — one (in Bonn) of the North Atlantic Treaty Organization (NATO) and another (at Versailles) of the members of the sub-set of the Organization for Economic Cooperation and Development (OECD) considered to deserve a seat at gatherings where senior politicians contemplate the future of the world. President François Mitterrand of France, who will not be at Bonn because France does not belong to NATO, has characteristically done his best to ensure that the Versailles summit, under the wing of OECD, will be enlivening. He has hatched a plan that the economic summit should begin with a discussion of governments' responsibilities for the development of technology. If he were not the host, the *rentier* in the public cause of the splendid palace at which the meeting will take place, it is improbable that his subject would have reached the head of the agenda. But the subject cannot be the emollient of fellow feeling that he calculated it to be.

President Mitterrand may in retrospect be thought to have captured the allegiance of those who voted for him a year ago by his simplicity; he seems to believe in what he says. This phenomenal habit appears also to be the reason why he has also won the affection, even respect, of European politicians ideologically at odds with him. But President Mitterrand also believes in the future — a future fashioned by technology. Why else would he have given such encouragement to his buoyant colleague, M. Jean-Pierre Chevènement, minister for research and technology? Together, they appear to be convinced that the future of France can be assured by the intelligent direction and the generous support of research and development. They may well be right. But President Mitterrand is also now bent on propagating this hopeful message in a wider circle, among France's industrial partners and competitors. The objectives may again be admirable, but it is important that President Mitterrand should understand why his advocacy next month at Versailles of the case that science and technology are a means of making up for economic recession will cause many of those present to shuffle uneasily in their seats. For the French argument, while half right, is three-quarters wrong.

The case that science and technology are too much neglected by those who look for a way out of present economic troubles is, of course, beyond dispute. If the present recession is caused by the fivefold increase of the world price of petroleum in the past decade, then something like a ten per cent increase of productivity in the oil-consuming states would suffice to ensure that nobody would notice. That, of course, is well within the immediate grasp of people in research and development. And a fivefold increase of productivity, by no means out of court, would ensure that the cost of buying oil would at some point in the future seem as unimportant as in the early 1960s, but with the difference that average prosperity would be five times greater. So, by putting money into research and development, the governments of industrialized states can realistically hope to work their way towards a more stable world, one in which not merely their electors but the people who live elsewhere will have a sense of being better off. Would that President Mitterrand's colleagues at next month's meeting at Versailles more openly shared his view.

The weakness of the French case is that it is almost certain to be an exaggeration. For at least the past two decades, successive French governments have worked deliberately at the improvement of science and technology. And not before time. The old republic was uncongenial for research. President de Gaulle, concerned though he may have been with the paramountcy of the identity of France, and of its place in the world, nevertheless provided an opportunity for those who sought to remedy the earlier neglect of the universities and of research. It tends, however, to be forgotten that even the young turks who took up that challenge made several disastrous mistakes. While the development of the petrochemicals industry in the early 1960s was an immense success, a proof that technology can spell prosperity, the nuclear power industry of that time was swiftly overtaken by events (and by the superior designs put on offer by United States manufacturers) while the efforts of those early governments to build a domestic computer industry were no more successful than later attempts have been. The still new government of France has inherited a much stronger base for the conduct of research and development than any of its predecessors; the universities are mostly in good shape, as are the grant-making agencies. But the government seems not to have learned from the hard experience of earlier decades. Its political persuasions, indeed, tell it that it need not learn. Yet mistakes will happen, and governments are especially prone to them.

This is why the unexpected French agenda item for the Versailles summit meeting will not be the uncontentious matter that President Mitterrand seems to hope. While some among the governments of the United Kingdom, the United States and West Germany may have ideological reasons for believing that too much government intervention in the shaping of the pattern of technology is mistaken, each of them also has vivid illustrations of how such efforts can go wrong. Sir Harold Wilson's technological revolution (1974) is only the most memorable error. Even the Japanese experience is not the proof that sufficiently clever governments can act wisely that the French will be hoping to claim at Versailles. For while the Tokyo government's support of successful manufacturers is the envy of manufacturers elsewhere (and properly a cause for complaint by luckless importers in other countries), nobody suggests that the government of Japan plans in advance whether next year's wonder product will be a new hi-fi set or a new space invaders game. The principle seems to be that the market will define what technology should do — if the market is given half a chance.

The French, in their present mood, are unconvinced. And it is certainly the case that if President Mitterrand could win the agreement of other industrialized nations to some coordinated plan for the development of technology, it would be possible to create both rising prosperity and a proof that the technological prophecy had come true. The agreement will not be forthcoming, but the proof would in any case be false. For to ordinary people, governments' electors, what matters about technology is not that it should be successful but that it should be wanted. Even the belief (to which the French government holds) that science and technology should be harnessed to the needs of development overseas is likely to run foul of the now well-established truth that developing countries do not like being told what is good for them. Sooner or later, Frenchmen will come to hold the same opinion if President Mitterrand pushes his view of the management of technology down their throats.

0028-0836/82/200170-01$01.00

US sours on foreign graduates

Job prospects clouded by immigration bill

Washington

US high technology industries and scientific organizations are just waking up to a little-noted, devastating change in US immigration laws which, if enacted, could change the face of Silicon Valley and university engineering faculties. The proposed change would prohibit foreign students from taking jobs in the United States on graduation. Instead, they would have to go home for two years before they could apply to work in the United States.

Under present law, they may apply for a change in visa status without leaving the country, and often do so.

The new provision may well become law before Congress adjourns. Identically worded provisions are in both the House and Senate versions of the bill. Neither has yet emerged from committee, but the provisions are unlikely to change before Congress acts.

Nobody knows how many students would be affected partly because many are either in the country illegally or do not report their work status.

At Stanford, for example, 50 per cent of all students studying for a PhD degree are foreign. Nationally, in 1980, 20.4 per cent of all science and engineering graduate students were foreign. In engineering, one-third of all PhD recipients are not US citizens. Engineering is the most popular field for the 311,000 foreign students, graduate and undergraduate, estimated to be in the United States (see Table).

Civilian high technology companies have come to rely increasingly on newly qualified engineers and scientists who are not US citizens, says Pat Hubbard of the American Electronics Association, because the US defence build-up has caused defence industries to hive off many of the American graduates. William Cagney of the National Foreign Trade Council which represents 650 major companies adds: "When DuPont goes looking for a PhD in chemical engineering, it more often than not hires a foreigner; there simply aren't that many Americans to go around. These people are vital to US industry." Both trade groups are opposing the proposed change in the law, as is the Semiconductor Industry Association, representing many Silicon Valley firms.

The provision in question is Section 212 of a bill that changes the Immigration and Nationality Act, the umbrella law that governs US immigration policies. Reform of the act has long been a public issue in the United States, where many thousands of Mexicans cross the southern border to work on ranches, illegally, and where the plight of the Cuban and Haitian refugees drew world attention to the problems of assimilating still more people from abroad. A blue-ribbon study was made by the Select Commission on Immigration, headed by Father Theodore Hesbergh of Notre Dame. The Reagan Administration proposed its own views of how to tighten US immigration laws, but the bill that has received the most support is that proposed by Romano L. Mazzoli (Democrat, Kentucky) in the House and Alan K. Simpson (Republican, Wyoming) in Senate. That part of the bill involving foreign students did not appear in Father Hesbergh's recommendations nor in the Administration bill, and so went unnoticed until a few weeks ago.

At present, most foreign students in the United States have visas allowing them to stay as long as they are studying, but they are not allowed to work. On graduation, they may apply for permanent resident status — a step towards citizenship — if they have a job offer. They usually remain in the United States (working) until this change occurs. A minority of foreign students are in the United States on so-called J visas, having been sent by their government or on exchange programmes, and they must return home on completion of their studies. Under Section 212, all students would have the equivalent of J status.

A House staff member explained that far

Setback for Chevènement

The second constitutional airing of the prospective science and technology bill (*loi*), on which the French science boom depends, has led to a bloodbath. For the Senate, the conservative French upper house, last week voted amendments which amputate most of its effective clauses.

The law, in the tattered form approved by the Senate, will now go to the National Assembly (the lower house), probably in June. Although the government has the majority there, the law will have to go through a lengthy process in which clauses deleted by Senate are reintroduced as amendments, adding to an already packed parliamentary session.

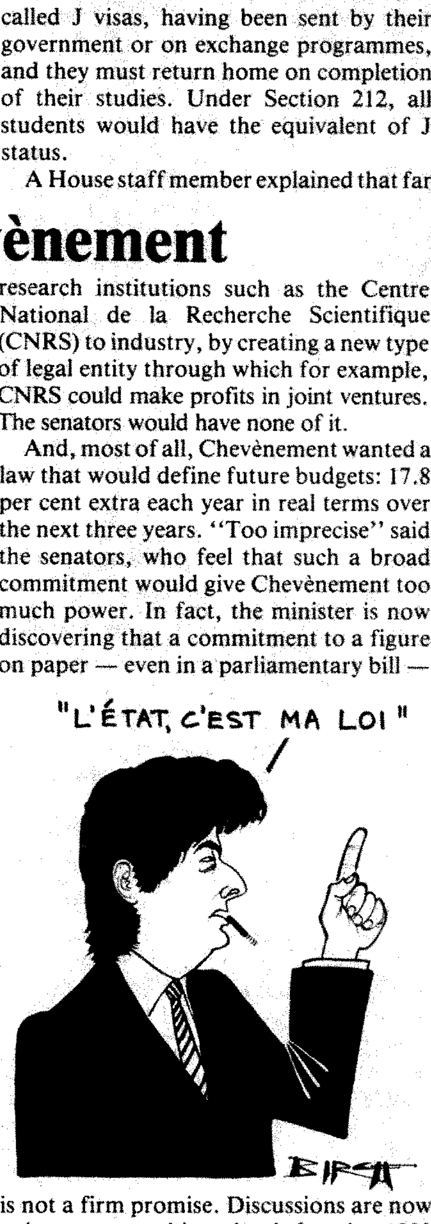
The Senate is an indirectly elected body, a third of its members being appointed each year on the basis of regional (departmental) elections. Electors consist of mayors and other such dignitaries, and amount to some 500 from each department. The Senate remains firmly conservative despite last year's election of the socialist Mitterrand government, and so is a source of constant irritation. But this is the first time it has behaved quite as badly towards government legislation.

The government explanation is that the science and technology bill gave the opposition an excellent opportunity to put a major dent in the plans of a senior minister, Jean-Pierre Chevènement, who is also the leader of the most left-wing element of the socialist party, the Ceres group. Ceres is of an intellectual and loosely Marxist tendency, and a favourite target of the conservatives. The fact that science and technology were involved in the Senate debacle is incidental.

Chevènement wished to devolve power to the regions, setting up regional councils for science and technology with their own budgets. The senators wished to "put off" that reform.

Chevènement wanted to open up research institutions such as the Centre National de la Recherche Scientifique (CNRS) to industry, by creating a new type of legal entity through which for example, CNRS could make profits in joint ventures. The senators would have none of it.

And, most of all, Chevènement wanted a law that would define future budgets: 17.8 per cent extra each year in real terms over the next three years. "Too imprecise" said the senators, who feel that such a broad commitment would give Chevènement too much power. In fact, the minister is now discovering that a commitment to a figure on paper — even in a parliamentary bill —



is not a firm promise. Discussions are now going on at cabinet level for the 1983 budget, and Chevènement is having to fight as hard this year as he did last.

Chevènement was impatient with the Senate's reaction to his well-laid plans. He described the Senate majority as "reactionary" and "cut off from the people". But he now faces a long struggle against the conservatives and against those of his cabinet colleagues who resent his scientific imperialism. **Robert Walgate**

0028-0836/82/200171-02$01.00

too many foreigners arrive in the United States saying they are going to study, but really plan to stay afterwards and work. This is a "backdoor immigration policy", he said, which must be reformed as part of a general tightening of US immigration.

The provision may have originated with an engineering activist, Irwin Feerst. Feerst says there was no consideration of a change in the old law regarding students, until he sent out a special issue of an independent newsletter he publishes to engineers around the country. Three hundred responses he received, he says, indicated that professional US engineers want to "throw the foreign students out". Feerst spoke to Senator Simpson and testified to this effect in December. According to Feerst, there is no shortage of engineers in the United States, only a "cabal" of academic engineers and corporate executives who are publicizing the alleged shortage so they can hire foreign graduates at lower pay.

Nationality of foreign students in the United States 1980–81

Regions	Selected sub-totals	Totals
Africa		38,180
Nigeria	17,350	
Europe		25,330
UK	4,440	
Greece	3,750	
FRG	3,310	
France	2,570	
Eastern Europe + USSR	1,670	
Latin America		49,810
Middle East		84,710
Iran	47,550	
Saudi Arabia	10,440	
North America		14,790
Oceania		4,180
South and East Asia		94,640
Taiwan	19,460	
Japan	13,500	
India	9,250	
		311,640

Source: Institute of International Education

However, the Institute of Electrical and Electronic Engineers (IEEE), the engineers' umbrella society in the United States (where Feerst has often run as an "alternative" candidate for president) does not favour the change in Section 212 as it is now written. Richard J. Gowen, chairman of IEEE's manpower task force and a candidate for president of IEEE, says that foreign engineers are being offered jobs at salaries that may be as little as 25 per cent of the salaries paid to US-born engineers. Gowen wants the proposed Section 212 to be changed to allow only "professional" hiring of foreign graduates. That is, a foreign graduate would have to go home unless his prospective employer certifies that he will be paid at least 75 per cent of what a US citizen would be paid in the job.

So far, Congress has been mainly concerned with getting the immigration reform bill through, and seems little disposed to tinker with the minor provisions. On the other hand, organizations like IEEE are becoming very active on the issue, as they begin to realize its implications. **Deborah Shapley**

EEC budget

Windfall ahead

Brussels

An unexpected shortfall of £280 million (500 million European Currency Units) in the European Economic Community's agricultural expenditure has resulted in an extra £19.6 million (35 million ECUs) becoming available for the European Community's research programmes in 1982. The strength of the dollar, favourable prices on the world market for agricultural goods and good weather have all helped to reduce the EEC's agricultural subsidies. Most of the extra money will be allocated to the joint research centres and in particular to nuclear safety research, but around £3.7 million will go to the indirect action programmes.

The budgetary revision highlights the degree to which the European Commission's ability to meet its goal of revitalizing scientific research and development in Europe, and thus create new jobs for scientists, is linked with the long-standing quarrel over agricultural prices, the British budgetary contribution, the reform of the Community's budget and the Common Agricultural Policy and even the Falklands crisis. If an agreement could be reached on lower agricultural prices for this year, as the British are demanding, an even greater sum could be set free for other areas of expenditure. The cooperation shown by the other member states over the Falklands crisis has now made it more difficult for Britain to push for budgetary reform. But if Britain achieves its objectives in the Council meetings this week, not only will its contributions be reduced, but it should lead to greater Community outlays in other areas including research.

The Commission's preliminary draft budget for 1983 reflects the hope that the other nine member states will agree to shift the emphasis of Community spending. Thus, the text emphasizes a "significant reinforcement of financial resources for energy policy, innovation and research and development". This includes increasing expenditure on energy research by 120 per cent although the total sum for payment appropriations will still be small, at £58.4 million.

Most of the extra money for this year will go towards the Supersara project on reactor safety at the joint research centre at Ispra in Italy. An extra £4.2 million will be needed this year as the project is overshooting its budget and will probably continue doing so until 1990 to the tune of £168 million.

The entire programme for the joint research centres is, in fact, now being reviewed and some changes are certain to be reflected in the 1984–87 research programme. Under consideration is a temporary increase in staff by 161, to provide replacements for scientists expected to retire in the next few years. Apart from streamlining research on nuclear safety, the handling of radioactive wastes and the control of fissile materials, a new institute for developing countries is planned at Ispra for training in energy planning, new energies, remote sensing techniques and the inventoring of resources. The Commission is also hoping to increase staff for research into solar energy, fusion, the rational use of energy and the study of high temperature materials.

Again, the success of these proposals will depend on the attitude of the member states, who will be influenced by the amount of money in the budget left over after provisions have been made for agricultural subsidies. **Jasper Becker**

Polish arrest

Expel and detain

The expulsion from Poland, last week, of two US diplomats and the arrest of Dr Ryszard Herczynski bodes ill for the resumption of normal academic exchanges between Poland and the West. The two diplomats, Scientific Attaché John William Zerolis and First Secretary for Cultural Affairs, James Daniel Howard, found in Dr Herczynski's flat, were accused by the Polish ministry of foreign affairs of "pursuing activity conflicting with their diplomatic status". Their presence in the flat appears, however, to have been entirely in the line of duty; they had gone there to confer with Dr Herczynski and with Professor Wladyslaw Fiszdon, a former pro-rector of Warsaw University, on the forthcoming joint US–Polish symposium on fluid mechanics.

Dr Herczynski, a mathematician specializing in fluid dynamics, is employed at the Polish Academy of Sciences' Institute for Fundamental Problems of Technology. Although now accused of having been "one of the inspirers of activity contrary to our *raison d'être* in the scientific milieu", he has never been associated with the dissident movement. In autumn 1980, however, he founded the "Society of the friends of science" — a semi-popular organization based on the then current principles of the liberalization of learning.

It was presumably for this reason that, during the night of 12–13 December 1981, he was taken into custody and interned for some two weeks. Although the authorities now claim that before being released, he signed an undertaking to cease such activities, Dr Herczynski's friends insist that not only did he never sign such an undertaking, but that at the time of his release there had been no mention of any such document.

The arrest of Dr Herczynski was, according to official sources, effected as he handed Mr Zerolis a packet of material "damaging to the interests of the Polish People's Republic", apparently com-

0028-0836/82/200172-02$01.00

prising a personal letter to his son, at present studying in the United States, an unidentified leaflet and an anonymous "Code of conduct during this testing time" addressed to Polish academics. Dr Herczynski now faces trial before a summary court. **Vera Rich**

NATO civil research

More wanted

Applicants for NATO (North Atlantic Treaty Organization) fellowships will have less chance of success this year than ever before. Applications have risen by 30 per cent, while the number of fellowships (about 800) remains constant. The NATO science committee will have to apply "new criteria" to make selections, a spokesman said last week. One possibility is that group applicants will be favoured over individuals.

NATO staff pinpoint four reasons for the increase in applications — financial difficulties among the 15 member countries, an increase in the number of papers now citing NATO as a supporting agency, a deliberate "willingness" on the part of NATO to expand the programme and a policy of greater visibility, including advertising in *Nature*.

The willingess to expand, however, is restricted to the NATO Science Committee, headed by Frenchman Professor Robert Chabbal (at present NATO's Assistant Secretary-General for Scientific and Environmental Affairs). The Civil Budget Committee, from which Chabbal draws his funds, is not so willing. To cope with the increase in applications for fellowships and for summer-school sponsorship, and for increased travel costs, the committee would have to increase its budget next year by 25–30 per cent in real terms, to $23–24 million. In fact, it may be lucky to get 15 per cent extra, just enough to cover the depreciation of the Belgian franc.

Pressing his case, Chabbal claimed last week that the NATO civil science programme (which completely avoids military research) is substantial and important. It accounts for half of all summer-school and training fellowships. Schools such as the Ettore Majorana at Erice, Sicily, and Les Houches in the Alps, get 60 per cent of their money from NATO, said Chabbal.

Meanwhile the committee will press ahead with new plans. It runs advisory panels which help to provide seed money for communications in new fields, largely by establishing "advanced workshops", and this year it will create two new panels: one on global transport mechanisms (in the atmosphere, ocean and mantle) and one on the selective activation of molecules (for example by laser). These panels would be expected to launch six workshops a year for a maximum of five years.

The committee is also experimenting with links between an industry in one country and a university in another, in a programme dubbed the "double jump". Finance will be *à la carte* — only interested countries need support it. So far only two such fellowships have been organized, but many more are planned — Dr Mario di Lullo, organizer of the "double jump" programme, believes that it will not run into the protectionist difficulties that sometimes face the European Commission — that one nation's industry does not reveal its secrets to nationals of another. **Robert Walgate**

British biotechnology

Out of the blue

In an unusual move, the British University Grants Committee is earmarking part of its annual budget to develop a specific topic — biotechnology. The committee plans that £800,000 will be spent in each of the next three academic years on fostering biotechnology in a handful of universities. The committee had previously been reluctant to earmark grants, preferring that universities should spend their income as they wished, relying on the research councils to encourage centres in particular topics by means of research grants.

The scale of the recent budget cuts seems to have prompted a change of heart. The committee, worried that universities may pare all their activities rather than cut them selectively, clearly hopes that earmarked grants will make the future pattern of university research more pointed. The £800,000 set aside for biotechnology will be taken from the money reserved for restructuring the reduced university system which in the next academic year (1982–83) will be £50 million. Most of that sum is expected to be spent on payments to redundant academics, leaving uncertain the amount available for fostering priorities.

So far, three centres — at University College London, the University of Birmingham and the University of Manchester Institute of Science and Technology — have been awarded annual grants of £100,000 each to develop biotechnology. Five other centres are expected to receive similar grants soon. The money will be paid as a separate item in each of the next three years, after which it will be incorporated in the recipients' recurrent grants.

The grants committee says that the recipients must decide for themselves how to spend their extra money. Nevertheless, it expects them to forge closer links with industry, chiefly by encouraging the process engineering side of biotechnology, to develop postgraduate rather than undergraduate courses and to appoint some permanent staff, thus fulfilling the recommendation of a Royal Society report which nearly two years ago called for twenty more university posts in biotechnology.

The research councils welcome the new grants, seeing no conflict between the grants committee's assessment of priority and their own. The Science and Engineering Research Council, in particular, welcomes the grants as a way of supporting staff and equipment which could not be paid for out of its research awards. **Judy Redfearn**

Development and drugs

More not less

The latest shot in the long-running battle between the pharmaceuticals industry and its detractors, in which the health problems of developing countries provide the battleground, has been fired by the Office of Health Economics (OHE). Despite its governmental sounding title and Whitehall address, the office is sponsored by the UK pharmaceuticals industry and its main task is to carry out research on the economic aspects of medical care. Its latest contribution, *Medicines, Health and the Poor World* by David Taylor, is a response to recent criticisms of the industry's marketing practices.

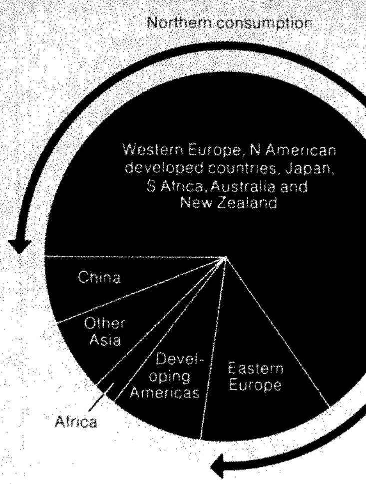
The large multinational companies have been accused of over-aggressively selling unsuitable drugs in the developing countries, leading to only a minimal improvement in the health of the population and sometimes proving positively harmful. Chief among the industry's critics have been aid organizations such as Oxfam and War on Want and the pressure group Social Audit.

The report acknowledges that some drugs have been "inappropriately" sold in the past but claims that the industry itself is now more capable of policing its methods of promotion and that the important role of drugs in improving health care in developing nations may be obscured by concentration on abuse in some areas.

Although a typical poor nation may spend around a quarter of its central government health budget on pharmaceuticals, the report says, 60–70 per cent of the people do not have regular access even to the most basic drugs. So while it is important that those drugs now being sold to the "wealthier" members of developing societies are properly advertised and correctly used, it is even more important to find ways of getting the basic drugs to the mass of the population deprived of them. Whether the past performance of the multinational companies has contributed to the weaknesses of governmental health services in the developing countries, or whether the unavoidable difficulties have limited the ability of the drug companies to act effectively, remains a point of conflict.

Better distribution of a limited range of medicines and vaccines, together with research aimed specifically at new pharmaceuticals for the developing world, are the urgent needs, says OHE. The World

0028-0836/82/200173-02$01.00



World pharmaceutical consumption for 1980 (total $76,000 million, manufacturer's prices.

Health Organization (WHO) has put considerable effort into work on new medicines and vaccines for diseases such as malaria, onchocerciasis, leprosy and leishmaniasis, but the report argues that the commercial companies cannot take on work on new drugs for these diseases because of the harsh economic realities of the marketplace — the return from sales of "Third World-oriented" drugs would not cover the cost of their development.

Thus OHE argues that the major drug companies should be used as contractors by WHO and governments to carry out the expensive first stages of developing drugs for diseases in developing countries. It is argued that this approach would be the most cost-effective way for the developed nations to help the poorer nations achieve better health standards.

To put the expenditure involved in finding new drugs into context, the total research and development spending by the UK pharmaceuticals industry is around £300 million per year, compared with the £1,000 million spent annually by the UK government on overseas aid. So an increase of just 5 per cent in this aid budget would provide £50 million, which if spent on contracted research could support a significant effort to find drugs important to developing countries.

On the supply of existing medicines to the developing countries, the OHE report is critical of the WHO programme on "essential" drugs. As for policing the marketing and advertising methods used to promote drugs in the Third World, OHE sees the international code of marketing practice drawn up recently by the International Federation of Pharmaceutical Manufacturing Associations as the best hope for the future. The federation, it is argued, would be well able to regulate the activities of the major international companies because it is in the companies' interest to be seen to be acting responsibly throughout the world. WHO is seen by the industry as being open to political pressures and therefore ineffectual as a controlling body.

Charles Wenz

US weather and ocean research

Outlook bleak

Washington

The major controversy in Washington this week has been the fate of the federal budget, as congressional leaders and the President joust with each other, leaving the outcome in doubt. No less concerned are the atmospheric scientists and oceanographers, especially those involved in international programmes, many of which have been cut back in the President's proposed budget. Ocean and atmospheric research has always had friends in Congress, and they are trying to have the funds restored. But whether the money can be put back, and then retained in a final budget package, is in doubt.

Marked for the axe are funds for joint US–Canadian efforts to clean up the Great Lakes, half of the satellite capability that provides weather data to nations in Asia, Africa and South America, funds to study and prevent ocean dumping (including radioactive waste), and the World Climate Program, the successor to the Global Atmospheric Research Program, which is a major international effort run from Geneva.

The Canadian government has been very upset about the proposed cut-back in US efforts to help study, monitor and clean up the Great Lakes, to which the US government is committed under several agreements with Canada. Under the proposed Reagan budget, two laboratories would be closed down and one programme office severely curtailed.

One laboratory is run by the Environmental Protection Agency (EPA) at Grosse Ile, Michigan, near Detroit. The other is run by the National Oceanic and Atmospheric Administration (NOAA) at Ann Arbor, Michigan.

The Administration justifies the cuts by claiming that most of the research needed to identify Great Lakes pollution has been done, and that cleaning up is a responsibility of the states that border them.

But the Canadian government's view is that still more pollution problems are being unearthed, such as the identification and reduction of such toxic substances as mirex and dioxin which have recently been found in "hot spots" in all the lakes except Lake Superior, and that the international commitments are federal, not state, issues. According to one Canadian official the cuts amount to "an attempt by the United States to renege on its water quality agreement of 1978" with Canada. This is strong language, given the historical close cooperation between the two countries — at least until the advent of the acid rain dispute and the Great Lakes budget cuts.

The Administration also wants to cut NOAA funds for ocean dumping research and marine pollution generally. NOAA runs most of the government's research in this area. The work concentrates on the area off the north-east coast of the United States which has the worst pollution problem, a major fishing ground and possible oil development. The rationale for the cut is that northeastern states should individually bear these costs. Congressmen seeking to restore the funds argue that these are national problems, especially in the light of renewed talk of disposing of radioactive and other forms of waste in the oceans.

For the second year in a row, the Administration has tried virtually to eliminate one of the ocean research programmes most popular on university campuses in the United States, the $35 million per year Sea Grant programme. For 1983, the Administration proposes only $1.7 million. However, as in 1982, representatives and senators are expected to get the funds restored. About half of the Sea Grant funds go for research, and the other half for community services related to the oceans issue.

The proposed Administration budget would also cut $6 million of the $7 million going to weather modification research, as part of a 40 per cent cut in atmospheric research funds. Likewise, a programme to upgrade the old weather radars on which continental US weather forecasting relies heavily is being slowed drastically. The Administration contends that weather is a local issue. Friends of NOAA counter, however, that weather modification and atmospheric research are important basic research programmes.

Of worldwide interest is a cut of $24 million that would decrease the launch rate of the NOAA series of polar-orbiting weather satellites, so that there would be one instead of two in orbit at any given time. These satellites complement the existing two GEOS geosynchronous satellites that provide "side views" of the Earth's weather, including the familiar television-screen images, to many countries that have receivers.

Representative James H. Scheuer, chairman of the House Science and Technology Committee's subcommittee on natural resources, agriculture research and environment, is trying to get the second NOAA satellite restored.

Scheuer argues that while the geosynchronous satellites indicate current weather, mostly in the middle and lower latitudes, the polar-orbiting satellites are essential to forecasting worldwide. Only they can acquire the quantitative data needed for modelling the Earth's weather, and only they can track the Arctic and Antarctic air masses moving towards the inhabited regions. The NOAA series satellites are thus crucial to weather prediction in parts of the world, such as Asia, Africa and Australia, that other satellites do not "look at" as often as at North America and Europe.

Two satellites provide twice as many passes over a region as one, and so improve

0028-0836/82/200174-02$01.00

daily forecasting. Two satellites also provide for redundancy. Scheuer says the Administration's cut will reduce the timeliness of forecasting and leave those parts of the world that are "neglected" by the rest of the system without any weather information at all when the remaining satellite breaks down — which happens often enough. Sometimes it takes a year to put a replacement for a broken satellite in orbit.

Finally, funds for the US contribution to the World Climate Program, a major international climatological effort run in cooperation with the World Meteorological Organization, have been cut from $1.8 million to $0.5 million.

Deborah Shapley

Lead in petrol

Clear risks

The British anti-lead lobby is flying high. Results now emerging from recent studies seem to be persuading scientific and medical opinion that the government should take further action on the lead content of petrol. Total elimination was advocated by many of the participants at an international conference organized last week by the Campaign for Lead Free Air (CLEAR).

Local councillors, civil servants and scientists from several countries heard evidence that children's IQ is impaired by low blood lead levels and that a significant contributor to lead burden is the lead in petrol. Des Wilson, chairman of CLEAR, is confident that the conference was given conclusive evidence that the position of the government is becoming increasingly isolated and untenable. The Lawther report, he said, had been completely discredited.

In 1980 a working party under the chairmanship of Professor P.J. Lawther advised that the evidence on danger from low blood lead levels was inconclusive. The Lawther working party also recommended that most effort be directed to reducing the lead in food and water, on the grounds that lead in petrol did not make the most significant contribution to the body lead burden.

The fear that blood lead concentrations of 300 $\mu g\ l^{-1}$ and below may have neurological effects represents a gradual shift of opinion over the past two years. The official government position is that harmful effects do not occur below 350 $\mu g\ l^{-1}$. The change of heart by the government which resulted in the decision last year to cut the legal amount of lead in petrol from 0.4 mg per litre to 0.15 mg per litre was in keeping with the Lawther case for a progressive reduction. Professor Lawther now says that the elimination of lead in petrol did not conflict with the findings of his working party and that "if there were damn all, nobody would be happier than me".

While there is no dispute about the toxic effects of lead at high levels, controversy persists about the level at which harmful effects occur. After the Lawther working party reported, two members, Dr Richard Lansdown and Dr William Yule of the Institute of Psychiatry and Great Ormond Street Hospital for Sick Children, produced results suggesting a link between blood lead levels and IQ performance in children of the London suburb of Greenwich and claimed a significant difference of IQ performance with blood lead above and below 120 $\mu g\ l^{-1}$ (*Devl Med. Child Neurol.* **23**, 567-576; 1981).

At last week's symposium, Yule and Lansdown gave new evidence (as yet unpublished) that children's behaviour is related to blood lead levels — 19 per cent of those with blood lead levels above the average (120 $\mu g\ l^{-1}$) were overactive compared with only 4.9 per cent of those with blood lead levels below the average. Bad conduct, nervous tension and lack of concentration increased with higher blood lead levels. Professor Herbert Needleman (Children's Hospital, Pittsburgh) also reported an association between classroom behaviour, IQ performance and lead burden.

Others say, however, that there should be extreme caution in moving from an observation of correlation between childhood lead exposure and impairment of intellectual development to postulating lead as the cause of extreme behavioural and psychological damage. The social environment is an extremely important factor in IQ performance.

The conference was also told of the significant contribution of lead in petrol to lead burden, making the Lawther estimate of 10 per cent seem too low. One recently completed study by the EEC Joint Research Centre at Ispra, Italy, suggests that as much as 30 per cent of the lead in blood is derived from petrol. The study determined the contribution of lead from petrol by relying on the known abnormal isotopic composition of lead in the petrol in one region of Italy. The conference also heard that in the United States in the four years since the phasing out of lead in petrol (1976–80), blood lead levels had fallen by 36.7 per cent. Dr Clair Patterson (California Institute of Technology) said that "exhausts from leaded gasolines are the most serious sources of lead in people".

While opinion is still divided and the need for further research acknowledged, Professor Michael Rutter (Institute of Psychiatry, London), a member of the Lawther committee, came down firmly in favour of the elimination of lead from petrol: "the level of probability is such that I think it is worth acting on".

CLEAR moves confidently into battle having already captured the support of the Labour party. The National Executive Committee pledged last month that a future Labour government would eliminate lead in petrol. CLEAR's sights are now set on the party conferences where it is confident of rallying support from the other opposition parties.

While much of the new evidence is not yet published, the Royal Commission on Environmental Pollution has now also launched an investigation into lead. It has heard evidence from the British Medical Association (BMA) supporting the link between low blood lead levels and impaired mental function. The BMA also appears to be convinced by the Ispra study.

The Royal Commission plans to investigate the sources of lead in the environment and its pathways to man. The commission also plans to "get to the bottom" of arguments about the technical and economic implications of reducing lead in petrol below that promised by the government last year. The commission may be confronted with strongly held opinions — Des Wilson described Associated Octel, the British company owned by Shell, BP, Chevron, Mobil and Texaco, which supplies the lead for petrol, as "the biggest mass child poisoners in the world today".

The oil companies say, however, that they will cooperate with the government if it calls for a ban on lead in petrol. The buck has now been passed.

Jane Wynn

Enter Exxon

New York

Cold Spring Harbor Laboratory on Long Island, New York, has agreed to a five-year "cooperative research agreement" with Exxon Research and Engineering Company. However, the financial terms of the deal are not being disclosed by Exxon, which is making its debut into the overcrowded world of biotechnology.

Under the terms of the agreement, up to six Exxon scientists will be assimilated into the Cold Spring Harbor staff to work full time on mutually agreed projects. In return, Cold Spring Harbor will select six postdoctoral fellows to participate in Exxon-funded research. In addition, Exxon's biosciences laboratory in New Jersey (which has never worked on molecular biology) may consult Cold Spring Harbor on various matters.

Exxon plans that its lawyers will visit the Cold Spring Harbor Laboratory regularly to keep abreast of latest developments. It will have exclusive rights to all patents derived from the research it funds, while Cold Spring Harbor Laboratory will retain rights to patents derived from work done by staff members not associated with Exxon projects.

Exxon will not ask those working on its Cold Spring Harbor Laboratory projects to defer publication of patentable inventions, as some industrial sponsors have requested of academic institutions relying on their funds.

Michael D. Stein

0028-0836/82/200175-01$01.00

CORRESPONDENCE

Scientists who act

SIR — Sir Peter Medawar's review of Martin Gardner's book "Science: Good, Bad and Bogus" (*Nature* 28 January, p.351), criticizes people working in fringe science, scientists who step aside from their specialisms to apply their experience in other areas and the inexperienced whose opinions force them into a corner where they grimly hold on to their untenable theories. He paints a broad canvas of condemnation and thereby implies that most speculation is put forward by rogues motivated by profit or starved of recognition and that we on the receiving end are unable to distinguish between the possible, probable and downright foolish.

Sir Peter's attack questions the integrity of scientists whose imaginative speculations, right or wrong, stimulate action that eventually leads to the truth. He makes use of the reference to unnamed people — the astronomer, microbiologists — without stating his objections. Above all, Sir Peter forgets that absence of proof is not proof of absence. Truth prevails eventually and the speculations and experiments that lead to it are winnowed by time.

Sir Peter's preoccupation with the intellectual underworld does disservice to those who think as well as labour in their laboratories. Also to those who in these difficult times are without laboratories and who have only their thoughts to offer to the service of human understanding.

D.G. APPLIN

Leystonstone,
London, UK

Doves in false garb

SIR — The leading article "Doves in false garb" (*Nature* 18 February, p.542) asks "What other justification is there for the Medical Campaign against Nuclear War except that physicians prefer hobnobbing with other physicians than with the hoi polloi?" This is a cheap shot unworthy of *Nature*. Let us hear some specific criticism of the physicians' proposals, let us hear some refutation of the plentiful justifications offered for the campaign by the physicians, but let us not hear "Nyah, nyah, nyah."

As the present American government considers a nuclear war ever more wageable and winnable, physicians must also begin ever more often to picture themselves crawling about on radioactive heaps attempting to dispense medicine in the aftermath of victory. The hope for curing anybody is less than small. Physicians claim that a nuclear war is a terminal disease for entire nations. When a doctor helps us to not catch a disease this is called preventative medicine. Such help is not only legitimate for a doctor to offer but may be felt required by the Hippocratic oath. Physicians involved in the aforementioned campaign are doing what they perceive is their duty to keep their patients alive. Does *Nature* doubt this prognosis? Do you share the opinion of the Federal Emergency Management Agency to the effect that society (at least that of the United States) will totally recover from a full-scale nuclear war within two to four years?

Nature concludes this commentary with the suggestion that if physicians, or any other professional group's members, seek to influence policy they should become members of parliament. This suggestion is nonsense. It is precisely because physicians are not politicians that they may help bring significant arms control into being. In past comment (*Nature* **294**, 197; **295**, 270) *Nature* has both stated the evident desirability of serious arms control negotiations and depicted the bumbling attempts of (US) politicians to initiate them. Atomic and hydrogen bombs now proliferate faster than ever before and members of parliament *et al.* have been as yet unable to inhibit this malignant growth. To stop and reverse the arms race will require the efforts of people in all walks of life acting in ways often without precedent, often compelled by personal, moral feeling.

Nature owes the living minds of its readers mature and honest evaluations of proposals from any campaign against nuclear war.

TOBIAS ISAAC BASKIN

Stanford Mid-Peninsula Chapter,
Physicians for Social Responsibility,
Stanford, California, USA

Mutualistic lives

SIR — In his News and Views article, "Mutualism: New Ecological Theories", May[1] asserts that ". . . conspicuous mutualistic associations tend to be tropical ones". May I suggest that conspicuousness is in the eye of the beholder? All the examples he cites involve animals. From ericaceous heaths of arctic–alpine regions, through vast acreages of boreal forest and temperate grasslands to tropical rain forests, the structure of the plant communities is governed, to those with the eye for them, by the conspicuous associations between plants and their mutualistic mycorrhizal fungi (see, for example, refs 2–6).

Mutualistic associations between plants and nitrogen-fixing prokaryotes and lichens are also important in tropical and non-tropical, especially sub-arctic, zones[7,8]. Harley[9,10], myself[11] and others have bemoaned the slow recognition by taxonomists and ecologists of the importance of mutualistic fungi and bacteria to plants in all terrestrial ecosystems. Let us hope that another of May's assertions[1] — that empirical and theoretical studies of mutualistic associations are likely to be one of the growth industries of the 1980s — will be world-wide and encompass interactions between all five kingdoms of organisms[12].

D.H. LEWIS

The University,
Sheffield, UK

1. May, R.M. *Nature* **296**, 803 (1982).
2. Harley, J.L. *The Biology of Mycorrhiza*, 2nd edn (Hill, London, 1969).
3. Janos, D.P. *Biotropica* **12**, 56 (1980).
4. Malloch, D.W., Pirozynski, K.A. & Raven, P.H. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2113 (1980).
5. Malajczuk, N. & Lamont, B.B. in *Ecosystems of the World, Vol.9B Heathlands and Related Shrublands: Analytical Studies* (ed. Specht, R.L.) 165 (Amsterdam, Elsevier, 1981).
6. Read, D.J. *Can. J. Bot.* (in the press).
7. Sprent, J.I. *The Biology of Nitrogen-fixing Organisms* (McGraw-Hill, London, 1979).
8. Hale, M.E. *The Biology of Lichens*, 2nd edn (Arnold, London, 1974).
9. Harley, J.L. *New Phytol.* **67**, 979 (1968).
10. Harley, J.L. in *Ecological Aspects of the Mineral Nutrition of Plants* (ed. Rorison, I.H.) 437 (Blackwell, Oxford, 1969).
11. Lewis, D.H. in *Taxonomy and Ecology* (ed. Heywood, V.H.) 151 (Academic, London, 1973).
12. Margulis, L. & Schwartz, K.V. *Five Kingdoms. An Illustrated Guide to the Phyla of Life on Earth* (Freeman, San Francisco, 1982).

Kin to whom?

SIR — Kin selection is not normally used by students of mammalian social evolution to explain the sociobiology of a particular class[1,2]. Contrary to the assertions by Packer and Pusey (*Nature* 22 April, p.740), Boorman and Levitt[3] (p.250) derived the conclusion that "attempts to apply $k > (1/r)$ directly to the analysis of primate, carnivore or other mammalian societies should be evaluated very cautiously" from their thorough review of the literature on mammals.

Students of vertebrate social systems have recognized the importance of the "multi-male" group, selection for outbreeding, and reciprocal associations between unrelated males for social evolution in the mammals, and the implications of these classical features for the genetic and phenotypic structures of mammalian, including human, populations have been discussed[1,4-9]. It is the persistence of these features that leads some primatologists to conclude that macaques and baboons, rather than gorillas or chimpanzees, are the richest models extant of hominid evolution[10,11]. Further, numerous "theoretical analyses" have pointed out that "inclusive-fitness maximizing" and kin selection are not equivalent and that under certain competitive regimes, kin may be "ego's" worst enemies[12-14]. Packer and Pusey's appreciation that individual reproductive "games" may not often be explained by kin selection for the mammals follows from the literature on behaviour and social evolution in that class.

CLARA B. JONES

Museum of Comparative Zoology,
Harvard University, Cambridge,
Massachusetts, USA

1. McCracken, G. & Bradbury, J. *Science* **198**, 303-306 (1977).
2. Schwartz, O.A. & Armitage, K.B. *Science* **207**, 665-667 (1980).
3. Boorman, S.A. & Levitt, P.R. *The Genetics of Altruism* (Academic, New York, 1980).
4. Trivers, R.L. *Q. Rev. Biol.* **46**, 35-57 (1971).
5. Eisenberg, J.F. *Handb. Zool.* **10**, 1-97 (1966).
6. Eisenberg, J.F., Muckenhirn, N.A. & Rudran, R. *Science* **176**, 863-874 (1972).
7. Alexander, R.D. *et al.* in *Evolutionary Biology and Human Social Behavior: An Anthropological Perspective* (eds Chagnon, N.A. & Irons, W.) 402-435 (Duxbury, North Scituate, Massachusetts, 1979).
8. Brown, J.L. *The Evolution of Behavior* (Norton, New York, 1975).
9. West-Eberhard, M.J. *Proc. Am. Phil. Soc.* **123**, 222-234 (1979).
10. Washburn, S.L. & DeVore, I. *Scient. Am.* **614**, 2-11 (1961).
11. Tiger, L. *Men in Groups* (Random House, New York, 1969).
12. Alexander, R.D. *A. Rev. Ecol. Syst.* **5**, 325-383 (1974).
13. Kurland, J.A. *Ethol. Sociobiol.* **1**, 255-274 (1980).
14. Waded, M.J. & Breden, F. *Behav. Ecol. Sociobiol.* **7**, 167-172 (1980).

NEWS AND VIEWS

The garden path

from Beverly Griffin

PERHAPS *Pilgrim's Progress* should after long neglect once again take its place on the literary scene, as required reading for today's molecular and cell biologists, and immunologists, instructing us to tread warily into the unknown, encumbered as it is with the snares, crevices and hidden dangers afforded by our modern technology. Not that there is anything sinister in the technology itself. Rather, we understand it too little and often use it too glibly. With due apologies to a distinguished group of scientists, I wish to quote a sentence from one of their papers to illustrate my point: "These results unequivocally demonstrate that chicken DNA contains nucleotide sequences related to (the) transforming gene of Y73 and that these endogenous DNA sequences . . . are different from the endogenous *sarc* related to *src* of RSV and do not correspond to any endogenous viral sequences"[1]. (Both Y73 and RSV, Rous sarcoma virus, are RNA viruses capable of inducing fibrosarcomas in chickens, and their transforming genes are designated *yes* and *src* respectively.)

The data that indicated *yes* and *src* as being structurally unrelated were based on nucleic acid hybridization analyses in conditions that would generally be described as 'medium stringency'. They were said to "unequivocally demonstrate" and were no doubt correct, but nonetheless led to incorrect conclusions, as now shown in an article in this issue of *Nature* (p.205) by N. Kitamura and some of the authors of the above mentioned paper.

Kitamura *et al.* have sequenced the *yes* gene of Y73 and compared it with a (revised) sequence of *src*. Overall, the homology is about 70 per cent; but they find that, at the nucleotide level, the longest homologous stretch is only 17 residues, which explains why in the conditions used[1], no hybridization was previously observed. A great many of the non-homologies, however, prove to be third base changes, so the encoded protein structures are remarkably similar, allowing the authors to suggest that these two transforming viral genes were probably derived from a common prototype sequence. The two protein sequences, although differing at the N- and C-termini, are almost indistinguishable over about 80 per cent of their length. Notably retained is a polypeptide stretch of over fifty amino acids containing the phosphorylated tyrosine residue that now appears to be a common feature of transforming genes of both RNA and DNA viruses.

In this context, another interesting paper by Stoner and Schlief[2] questions whether the amino acid, but not the nucleotide, sequences of the *araC* genes in two strains of *Escherichia coli* have been conserved. Again, the data point to third base changes which would prevent hybridization but retain protein structure. Returning to the viral model, one can ask whether all the transforming genes of the class of viruses called retroviruses, when subjected to the same degree of scrutiny, might not prove to be similarly related. Take the *fps* gene in Fujinami virus (FSV), for example. Hanafusa *et al.*[3] state that "the results described convincingly exclude the presence of RSV-related *src* gene in the genome of FSV. The lack of *src* was shown by oligonucleotide fingerprinting, molecular hybridization and immunoprecipitation with antiserum to the *src* gene product". The question now arises, will these techniques survive the nucleotide sequence test, or have they, in this instance, proved false in the hands of undeniably eminent scientists? At this stage, the possible relationship between viral transforming genes and oncogenes[4,5] or that between the transforming genes of RNA and DNA tumour viruses seems a matter of conjecture. The question has to be raised of how much confidence one can put in negative data. With all the enthusiastic computing going on at many terminals, can a program be devised that will allow the accuracy of sequence relationships to be estimated, particularly with respect to hybridization?

Before this becomes possible, one presumably needs to understand what hybridization measures. Is it, for example, 90 per cent hydrogen bonding (G-C carrying more weight than A-T) and 10 per cent pi-electron interaction (purine–purine interactions more significant than pyrimidine–pyrimidine)? What effect does the location of imperfections in homologies have on the hybridization? How effectively does intramolecular hybridization compete with intermolecular hybridization and under what conditions? Since much weight is put on the data derived from hybridization between partly complementary sequences, shouldn't we know the answer to questions such as these? The future clearly will involve searching for specific cellular genes with the aid of this very powerful, but also less than well understood, technique.

The literature contains a smattering of good papers on the subject. To quote but a few that help to clarify matters, a start can be made with a paper by Howley *et al.*[6]. The authors chose three related and well defined (that is, sequenced) DNA viruses of different host origins to compare by hybridization — SV40 from monkey, BK virus from man and polyoma virus from mouse cells. Under stringent hybridization conditions, very little homology was observed, and this was assigned to the region encoding the viral capsid proteins, not a region of primary importance in the pathology of these viruses. By sequentially decreasing the stringency, however, homologies began to appear elsewhere, such that the coding regions of SV40 and BKV could be observed to be remarkably similar, and also similar to about 75 per cent of polyoma virus. The validity of these hybridization studies was confirmed by a direct comparison of their DNA sequences[7], and evolutionary relationships among the viruses were proposed[6,8]. What was the outcome of this less qualitative and rather more detailed approach to hybridization than is usually seen in the literature?

Beverly Griffin is in the Imperial Cancer Research Fund Laboratories, Lincoln's Inn Fields, London WC2A 3PX.

0028-0836/82/200177-02$01.00

One outcome was to suggest that immunological relationships should exist among some of the viral proteins other than the structural proteins. McCormick *et al.*[9] thus looked for, and (in contrast to previous negative reports) found, immunological cross-reaction between the large T(tumour) antigens of SV40 and polyoma virus. They conclude their paper with the warning note, "that a cross-reaction between two proteins of related function and extensively homologous primary sequences should prove so elusive sounds a note of caution in the interpretation of negative results when immunological techniques are used to search for homologies between polypeptides".

At a simpler level, but perhaps the one from which quantitative understanding of hybridization will ultimately come, Gillam *et al.*[10] have studied melting temperatures between duplexes of oligo(T) and oligomers such as A_5TA and A_5GA. A difference of several degrees in the relative stabilities of the duplexes was found, with the rather surprising result that the former mismatched double helix appeared to be more stable than the latter one. As the length of the oligomers was increased from the septamer to the nonamer, this difference diminished. (Among many interesting conclusions, they note that an oligodeoxyribonucleotide is less likely to form a mismatched structure with DNA than with RNA.) When homologies between dissimilar but related species are compared, is not hybridization essentially a result of a composite of melting out and reforming short duplexes?

A related approach has been taken in a series of stimulating papers by Wallace *et al.*[11], who investigated hybridization by using oligodeoxynucleotides as 'hybridization probes' for specific genes. By modification of the conditions under which hybridization was carried out, they could discriminate between the formation of perfectly matched and single-mismatched duplexes involving rabbit β-globin DNA and fourteen-long oligomers. To end on another 'perhaps', studies such as the elegant ones mentioned above may ultimately lead to a series of rules for the selection of the appropriate hybridization conditions that will allow interactions, whether between nucleic acids, or proteins, or even between the two, to be predicted accurately, even before detailed sequences become known. □

1. Yoshida, M., Kawai, S. & Toyoshima, K. *Nature* **287**, 653 (1980).
2. Stoner, C.M. & Schlief, R. *J. molec. Biol.* **154**, 649 (1982).
3. Hanafusa, T. *et. al. Proc. natn. Acad. Sci. U.S.A.* **77**, 3009 (1980).
4. Weinberg, R.A. *Trends biochem. Sci.* **135**, April (1982).
5. Bishop, J.M. *Scient. Am.* **69**, March (1982).
6. Howley, P.M. *et al. J. biol. Chem.* **254**, 4876 (1979).
7. Soeda, E. *et al. Nature* **283**, 445 (1980).
8. Soeda, E. *et al. Nature* **285**, 165 (1980).
9. McCormick, F., Lane, D.P. & Dilworth, S.M. *Virology* **116**, 382 (1982).
10. Gillam, S., Waterman, K. & Smith, M. *Nucleic Acids Res.* **2**, 625 (1975).
11. Wallace, R.B. *et al. Nucleic Acids Res.* **9**, 879 (1981), and references therein.

Fission, fusion and fracture

from Robert W. Cahn

CONFIDENCE in the intrinsic feasibility of an economically effective fusion reactor is now sufficiently widespread among those responsible for research policy that technological aspects of fusion reactor design are receiving detailed attention. One straw in the wind is the recent publication of the Proceedings of the Second Topical Meeting on Fusion Reactor Materials[1].

While about a quarter of the papers deal with such matters as tritium and deuterium recycling, breeding blankets and their constituent materials (especially lithium), and alloys for magnet windings, the central concern is the 'first wall' — the inner surface of the torus which is exposed to the full intensity of the 14 MeV neutrons resulting from the D-T fusion reaction. Schiller[2] surveys the various dangers that beset a first wall: swelling and embrittlement due mainly to helium generated *in situ* by a nuclear reaction due to neutrons, especially with nickel; erosion by sputtering due to neutral atoms escaping from the plasma, which might remove as much as 1cm from the surface during a 15-year reactor lifetime; and thermal fatigue because the operation of the reactor will be pulsed.

A number of alloys have been proposed for the first wall, including some exotic ones. Thus several investigators have seriously proposed the use of metallic glasses which typically crystallize at 350–600°C: Emmoth *et al.*[3] have shown that the familiar glass $Fe_{40}Ni_{40}B_{20}$ sputters three times more slowly than Fe-Ni-Cr (crystalline) stainless steel. (This is interesting from a fundamental point of view also since no rigorous comparative measurements have yet been made of sputtering yields from the same alloy in glassy and crystalline forms.)

However, the 'prime candidate materials' for the first wall appear to be an Al-Mg-Si alloy and variants of type 316 stainless steel (with ∿ 15 wt per cent each of Ni and Cr). A great deal of research on 316 steel was reported at Seattle: it is a familiar material, used for components of fast fission reactors as well. The most impressive of the papers on 316 steel variants are a group devoted to steels incorporating a fraction of a per cent of titanium carbide. One paper[4] from the very active research group at Jülich in West Germany compares the behaviour of 316 steel with and without TiC. The unirradiated creep rupture life at constant stress at 700°C and the reduction of creep rupture life due to helium ion implantation (which simulates what happens in a fusion reactor first wall) were very much better in the modified steel. This is plainly due to a fine distribution of TiC in grain interiors and the ability of these TiC particles to trap helium bubbles away from the grain boundaries where alone they produce creep embrittlement. This is an example of the increasingly common metallurgical ploy of setting one impurity trap to combat the effects of another.

A steel almost identical to that studied at Jülich is the American 'prime candidate path A alloy'. Arnberg *et al.*[5], of MIT, show that if this alloy is made by rapid solidification processing (flakes frozen at ∿ 10^5 K s^{-1} and then hot extruded), a fine grain size (∿ 10 μm) and extremely fine TiC particles result. This material, when helium-bombarded, proved to be very resistant to swelling at ∿ 600°C, even more so than the same alloy made by conventional methods. The TiC, according to the investigators' electron micrographs, nucleates numerous helium bubbles (especially the more numerous TiC particles in the rapidly frozen alloy) and these bubbles all remain very small, presumably under high pressure. Megusar *et al.*[6] (also at MIT) have prepared versions of this steel with triple the TiC concentration (only rapid freezing will give a uniform distribution of particles at this concentration) and this is expected to be even more resistant to swelling. The corresponding embrittlement tests have not yet been undertaken.

The 316 steels are austenitic (face-centre cubic in crystal structure). Ferritic (body-centred cubic) Fe-Cr steels without nickel have also been examined They have the advantage of generating less helium from (n,α) reactions than do 316 steels. These steels are hardened but also quite severely embrittled by irradiation[7], but Japanese work showed that a high Cr content (15 wt per cent) minimizes embrittlement[8].

An interesting German paper[9] shows that ordinary pressure vessel steel (without Ni or Cr) embrittles much less when irradiated at the service temperature of 290°C (in a PWR reactor pressure vessel) than when irradiated at ambient temperature, in the sense that the ductile-brittle transition temperature is less affected. However, if the steel contains a small amount of copper (a notorious source of contamination from recycled scrap steel), then embrittlement at service temperature is much more substantial. Here, an involuntary trace element, unlike the intentionally added Ti in 316 steel, has dangerous effects; but it is not difficult to control copper levels in high-grade steel.

Pressure vessel steels are currently being analysed in great detail in terms of linear elastic fracture mechanics, and the characterization of such steels goes a long way beyond a study of the effect of

Robert W. Cahn is Professor of Metallurgy at the University of Paris-Sud, Orsay 91405.

irradiation on ductile–brittle transition temperatures. It is possible, with a very high degree of confidence, to calculate for any part of a pressure vessel to what size a crack must grow before it can present any danger. The basis for these calculations have been outlined in technical terms by Nichols and Cowan[10] and, with his incomparable clarity of exposition, by Cottrell in his popular book on reactor safety[11]. While a failure of a fusion reactor first wall would certainly represent a far less serious accident than the fracture of a PWR pressure vessel (which, of course, has never happened), no doubt fracture mechanics will in due course be applied to the fusion situation as well, since it represents the guarantor of nuclear safety where the prevention of fracture is concerned. □

1. Proc. Second Topical Meeting on Fusion Reactor Materials, Seattle, August 1981 *J. nucl. Mater.* **103 & 104** (1981-2).
2. Schiller, P. *J. nucl. Mater.* **103/4,** 75 (1981).
3. Emmoth, B. *et al. J. nucl. Mater.* **103/4,** 393 (1981).
4. Kesternich, W. & Rothaut, J. *J. nucl. Mater.* **103/4,** 845 (1981).
5. Arnberg, L., Vander Sande, J.B., Frost, H.J. & Harling, O.L. *J. nucl. Mater.* **103/4,** 1069 (1981).
6. Megusar, J., Arnberg, L., Vander Sande, J.B. & Grant, N.J. *J. nucl. Mater.* **103/4,** 1103 (1981).
7. Kßueh, R.L., Vitek, J.M. & Grossbeck, M.L. *J.nucl. Mater.* **103/4,** 887 (1981).
8. Suganuma, K., Kayano, H. & Yajima, S. *J. nucl. Mater.* **105,** 23 (1982).
9. Ahlf, J. & Schmitt, F.J. *J. nucl. Mater.* **105,** 48 (1982).
10. Nichols, R.W. & Cowan, A. *Phil. Trans. R.Soc.* **A299,** 227 (1981).
11. Cottrell, A *How Safe is Nuclear Energy?* 60 (Heinemann, London, 1981); see also *Nature* **293,** 763 (1981).

Star formation in galactic spiral arms

from Gerard Gilmore

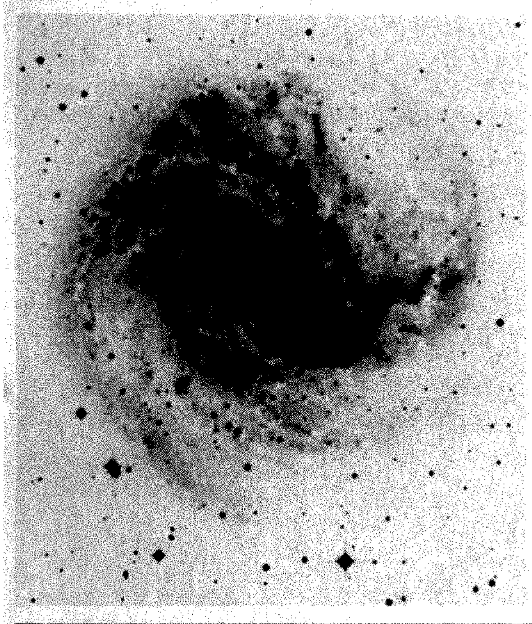
TWO-THIRDS of all bright galaxies show spiral structure, with the arms being made visible by young luminous stars and the giant gas clouds which are the sites of continuing star formation. The relation between the existence of the arms and the young stars which delineate them is analogous to the relation of the chicken to the egg. Do the galactic density waves which sweep material into the arms purely by chance create the ideal conditions for star formation, and thereby provide an optical tracer of the density enhancement, or is star formation a more fundamental requirement for the existence of the spiral arms? While current theories of galactic spiral structure suggest that the star formation is incidental, a recent numerical study of the effect of star formation on the appearance of a rotating galactic disk suggests that, in some cases at least, bursts of star formation may in fact generate the observed structure (Seiden, Schulman & Feitzinger *Astrophys. J.* **253**, 91; 1982).

Gerard Gilmore is at the Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ.

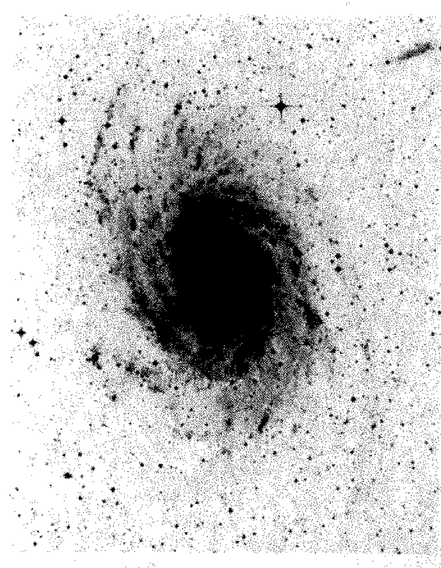
There are currently two widely discussed theories of galactic spiral structure. The first, originated by Lindblad in 1941 and developed extensively by C.C. Lin and associates, involves the response of a disk to an imposed spiral perturbation to the galactic gravitational field. As the stars and gas respond to this perturbation, they form a spiral density perturbation, which sustains the gravitational pertubation. The density wave is therefore self sustaining. The important difficulties with this theory are to find an origin for the perturbation, to maintain it against energy loss to resonances in the stellar orbits and to prevent its destruction by winding up. In general, the stars and gas outside the nuclear regions of galaxies have flat rotation curves, implying strongly differential rotation with distance from the galactic centre. This in turn will smear out any material structure in a few rotation periods. Spiral galaxies are, however, typically 50 rotation periods old. So what maintains the spiral structure?

In the second theory, extensively developed by Alar Toomre and collaborators, the spiral density perturbation is generated as a consequence of dynamical interactions with nearby companion galaxies. His numerical simulations of such interactions show a strong density wave propagating through the disk galaxy as a consequence of internal dynamical resonances (called 'swing amplification' by Toomre) initiated by quite small density perturbations. This model can generate density waves, but requires them to be transient. The high proportion of disk galaxies showing spiral structure is then a consequence of the high rate of dynamical interactions between galaxies. The spiral structure of our Galaxy becomes a consequence of the orbital properties of the other galaxies in the local group.

While proponents of both these theories are able to generate models in striking agreement with observation, what of isolated galaxies, where no obvious origin for the spiral density perturbation is apparent? It is here that the model of Seiden *et al.* is of interest. They have extended earlier work by Seiden and Gerola (*Astrophys. J.* **233**, 56; 1979; *Nature, News and Views* **283**, 133; 1980) modelling the effect of star formation on the appearance of a galactic disk by including a simple approximation to current ideas of self-propagating star formation (SPSF). In this model, a few very massive stars formed in a dense molecular cloud evolve rapidly, and generate strong shock waves from stellar winds or supernovae explosions. As these shocks propagate through the remaining dense cloud, they initiate further compression with consequent enhanced star formation, leading to more massive



Left, M83, a large, almost face-on spiral galaxy perhaps 10 million light-years away. The complex spiral arms are patterned with masses of dust, some of which seem to occur in giant shells many hundreds of light years across, and contain large numbers of young, luminous stars and masses of hot, glowing gas. At the centre of the galaxy is a source of radio waves. Right, one of the largest barred spiral galaxies, of visual magnitude 10.6 and an angular diameter of 9 arc min, in the constellation of Pavo.



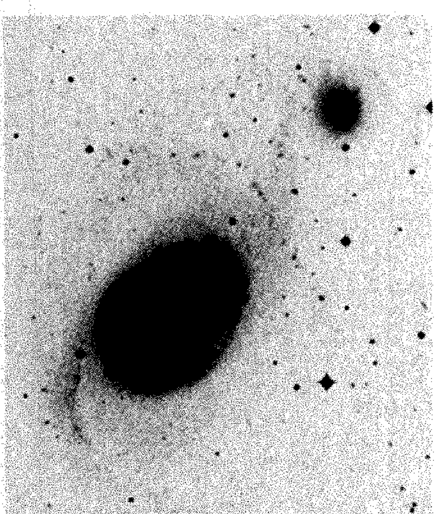
Photography by Photolabs, Royal Observatory, Edinburgh

0028-0836/82/200179-02$01.00

stars, and so on (Lada, Blitz & Elmegreen in *Protostars and Planets* ed. Gehrels, 341; 1978). The positive feedback rapidly destroys the conditions for further star formation, leaving a proportion of the cloud as gas. Eventually, this will cool and recollapse, ready for another burst of star formation.

Seiden *et al.* model this SPSF process by allowing random star formation to initiate bursts of subsequent formation, and explicitly including a cooling time for the remaining gas. As a burst of star formation propagates outwards, differential galactic rotation smears the expanding shell into an arc. The arc will then grow until it meets a region in which star formation occurred less than the recovery time before. With this simple view, the simulations are able to generate two-armed spiral galaxies allowing the initial star formation to occur randomly (hence their name: stochastic self-propagating star formation — SSPSF), and without having to impose an initial spiral density perturbation. They note that their process may in fact be an appropriate mechanism to stabilize a spiral density wave for many galactic rotation periods.

Gerola *et al.* (*Astrophys. J.* **242**, 517; 1980) have also pointed out a further important property of models of this type for small galaxies. If the recovery time for the gas is comparable to the galactic rotation period, the star formation will occur in delayed but periodic bursts. Large bursts of star formation have in fact been invoked by many authors for several years to explain the properties of small irregular galaxies and anomalously blue galaxies. While most of these bursts can be explained, as with spiral arms, by dynamical interactions of nearby companions, a few cannot. As Seiden *et al.* point out, it is not clear how to distinguish observationally between periodic bursts of star formation for the few isolated galaxies, as predicted by their model, and a non-periodic phenomenon.



The larger of the two galaxies is NGC1512, consisting of a cigar-shaped bar of stars surrounded by a bright ring from which the spiral arms trail. The small elliptical companion, NGC1510, contains many young stars and clouds of hot gas, features which are very rare in galaxies of this type. It is thought to have 'rejuvenated' itself by scooping up gas from the cloud around its larger companion; the newly acquired material has since formed stars.

Perhaps the best summary of the current state of spiral density wave theory is by Toomre (see *Structure and Evolution of Normal Galaxies* eds Fall & Bell, 111; 1981): "It seems clear now that the spiral structure of galaxies is a complex riddle, without any unique and tidy answer".□

The great Japanese IQ increase

from Alun M. Anderson

THE mean IQ of the Japanese has risen spectacularly this century and, among the younger generation, is now the highest in the world, some 11 points above the mean IQ of the United States and other Western populations, according to a report by Richard Lynn in this issue of *Nature* (p.222). Analysis of results from the American Wechsler Intelligence Scale for Children shows that more than three quarters of the Japanese younger generation have a higher IQ than the average American or European. Lynn points out that with this difference in mean IQ there will be a far greater proportion of the Japanese population with a high IQ — 10 per cent of the Japanese younger generation should have IQs above 130, the levels generally found among professional groups such as research scientists, engineers, lawyers and doctors, compared with only 2 per cent of the American population. If, indeed, any link between IQ and intellectual achievment is accepted then the future implication for societies which are becoming increasingly dependent upon technological innovation may be profound.

How can this remarkable increase be explained? The dramatic changes in Japanese society that have occurred in the last century — a mixing of previously isolated peasant communities in the massive post-war urbanization, rapid economic growth and accompanying improvements in welfare, health and education and exposure to Western culture and ways of thought — probably provide a large part of the explanation. By comparing the group of Japanese born in 1910–1945 (mean IQ 102–105, only a little above the American level) with that born between 1946 and 1969 (mean IQ 108–115) it is possible to find evidence of changes in that period in almost every factor, genetic, nutritional, educational and social, that has been supposed to affect IQ. Some of these can be documented by a few key measures of Japanese society.

Japan retained a feudal society much later than the West. Until 1868, when the Emperor was restored to power after nearly three centuries of rule by the military Shogunate, the great mass of the population were peasant farmers, tied to the land and marrying within the local

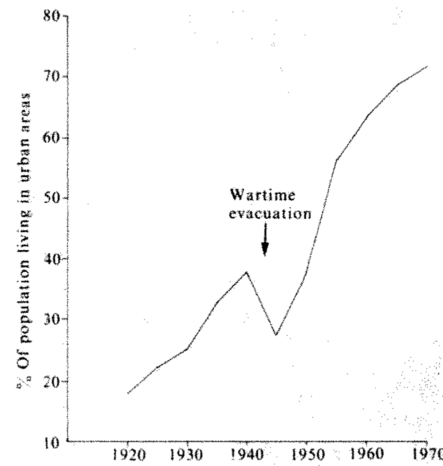


Fig.1 Increase in the urban population of Japan between 1920 and 1970 (from ref.1).

communities. In the late nineteenth and twentieth century a movement of the population towards the cities began, but it was not until the middle of this century that urbanization gained pace. In the period between 1930 and 1960 almost 40 per cent of the population moved from the country to the cities[1] (see Fig. 1). The remoter rural areas of Tohoku, Hokuriku, San'in, Shikoku and Kyushu were drained of their population, whole villages sometimes being abandoned, as Tokyo, Nagoya and Osaka emerged as massive industrial centres. In the new urban areas rural groups that had been isolated for many centuries now mixed and inter-marriage between individuals from different parts of the country became common. As a result the Japanese population has almost certainly become more outbred — a supposition supported by data on consanguineous marriages gathered in the exhaustive genetic studies of the populations of Hiroshima and Nagasaki carried out by the Atomic Bomb Casualty Commission[2]. The change is one factor that could have contributed towards the IQ increase; that inbreeding can reduce cognitive performance, including IQ scores, is well documented (see ref. 3, for example).

Raised nutritional standards are very likely to have had major effect on IQ. Lynn

Alun M. Anderson coordinates the News and Views columns of Nature.

0028-0836/82/200180-02$01.00

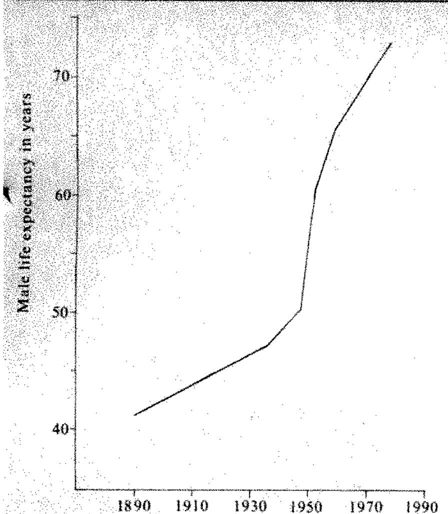


Fig.2 Male life expectancy in Japan between 1890 and 1977 (from ref.8).

points out their importance and documents them by noting the increase in birth weights. His conclusion is supported by the recent increases in the height of the Japanese and the rapid rise in life expectancy.

Japanese males born in 1960 are on average 5.2cm (2.1 inches) taller than those born in 1940. A similar increase was observed in Western Europe in the late 19th and early 20th century[4] and was closely linked to improving economic conditions. Interestingly, the change is not simply one of height but of proportion, for leg length has increased by 4.3 cm while sitting height (from base of spine to top of head) has increased by only 0.9cm in the same period[5]. (Thus the popular, but erroneous, Japanese view that they have grown taller by adopting the habit of sitting on Western chairs rather than sitting with legs folded underneath the body on a *tatami* mat.)

The rapid rise in life expectancy (to 72.7 years for men and 78.9 years for women in 1977 compared to 46.9 years and 48.6 years in 1935, see fig. 2) has produced yet another first for the Japanese. In 1977 they became the world's longest lived race and their life expectancy has continued to rise since then.

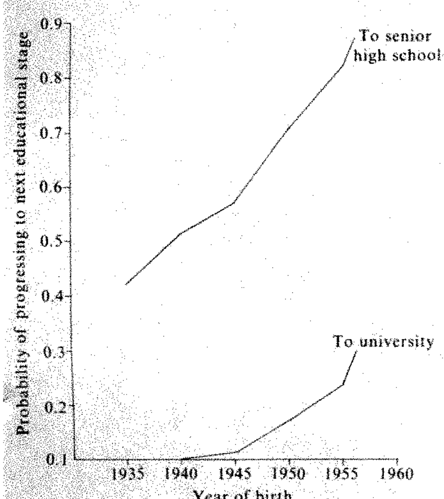


Fig.3 Fraction of middle school students going on to senior high school and to university or higher education college according to year of birth (from ref.9).

Of more obvious significance for the rise in IQ may be the very great increase in educational opportunities since the 1930s. School attendance is well known to have a substantial effect on IQ levels[6]. Among those born in 1948 and graduating from junior high school in 1962, only 51.5 per cent went on to senior high school and 10.3 per cent to college. By 1972, 87 per cent of junior high school students were progressing to senior high school and 30 per cent going on to college (see Fig. 3).

Nor, indeed, is this the first time that such a rise in IQ has been recorded. In the early days of American IQ testing many new immigrant groups were found to have low IQ (admittedly on much less sophisticated tests) but as they became accustomed to American life their IQs rose[7]. A similar effect may have occurred as the Japanese have become used to tests devised for a very different culture.

The change in IQ may, then, be explained in a multitude of ways. What remains a puzzle is why the rise should have been to a level significantly higher than that found today in the United States. Once again, given the massive cultural differences between the West and Japan and the difficulties of devising any test that does not favour one culture or another there would seem to be endless opportunity for speculation. Whether the difference in IQ represents a real difference in 'intelligence' or simply implies that the Japanese are better at IQ tests, remains open to question. At present, there is little more to say than that. □

1. Bureau of Statistics, Office of the Prime Minister, 1970 Population census of Japan (1971).
2. Schull, W.J. & Neel, J.V. *The Effects of Inbreeding on Japanese Children* (Harper & Row, New York, 1965).
3. Vogel, F. & Motulsky, A.G. *Human Genetics*, 428 (Springer Berlin, 1979).
4. Vogel, F. & Motulsky, A.G. *Human Genetics* 185 (Springer, Berlin, 1979).
5. Ministry of Education Survey (1977).
6. Jensen, A.R. *Straight Talk about Mental Tests*, 183 (The Free Press, 1981).
7. Gould, S.J. *The Mismeasure of Man* (Norton, 1981).
8. White Paper on Health and Welfare, Ministry of Health and Welfare (1977).
9. White Paper on Education, Ministry of Education (1977).

The 'fold-back' elements of *Drosophila*

from D.J. Finnegan

GENOMIC rearrangements and DNA transposition have been observed in eukaryotes so often as to be a novelty no longer. Their significance and the various mechanisms by which they occur are far from clear, however, and at this stage all examples bear detailed investigation. The best characterized eukaryotic transposable elements are the '*copia*-like' elements[1-3] of *Drosophila melanogaster*. They are not the only transposable elements in *D. melanogaster* and in this issue of *Nature* (p.201), Potter presents the complete DNA sequence of a member of a striking different class of transposable elements — the 'fold-back' elements.

When DNA from any eukaryote is denatured and allowed to re-anneal at low concentration the first sequences to become double-stranded are those that occur as inverted repeats. The repeats may be immediately adjacent or separated by up to several thousand base pairs and, as the reaction is intramolecular, the structures formed are called 'snap-back' or 'fold-back' structures. Schmidt *et al.*[4] and Baker and Thomas[5] have estimated that there are 2,000–4,000 pairs of inverted repeats, or potential fold-back structures, in *D. melanogaster* which comprise about three per cent of the total genome. To study the properties of the fold-back structures Potter *et al.*[6] screened a library of cloned fragments of *D. melanogaster* DNA with a probe enriched for the stems of fold-back structures. They obtained one clone containing an inverted repeat and used it to select a number of others containing related fold-back (FB) elements. The sequence of one of these, FB4, is reported in this issue.

Potter *et al.*[6] also determined the chromosomal distribution of FB elements by hybridizing an FB probe to polytene chromosomes of larvae from different strains. About 30 hybridization sites were found in each case but their actual positions differed, indicating that FB sequences can move about the genome, that is, that they are transposable elements.

There are approximately 25 families of *copia*-like elements in the *D. melanogaster* genome. The members of each family are well conserved and are bounded by direct repeat sequences a few hundred nucleotides long, at the ends of which are short inverted repeats. The sequence of the direct repeats varies little within a family and, in all cases looked at so far, the repeats at the end of any one element are identical. In contrast, FB elements vary greatly in their inverted repeat sequences and the DNA separating them[6,7]. The inverted repeats have a sequence organization similar to that of satellite DNA[7] and those of FB4 are made up largely of tandemly repeated copies of a 155 base pair repeat unit which itself contains smaller sub-repeats[8]. Inverted repeats of different FB elements vary in length from a few hundred to a few

D.J. Finnegan is in the Department of Molecular Biology, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JR.

0028-0836/82/200181-02$01.00

thousand base pairs, presumably reflecting variation in the number of repeat units present[7]. The inverted repeats of FB4 differ in this way, with the right-hand repeat being longer than the left, and Potter points out that length variation of this type could be generated by mismatch recombination between repeats of different FB elements. In addition to these gross differences, the two repeats also differ by base substitutions and deletions/additions. The most highly conserved regions are at the extreme ends of FB elements. This is not surprising since these regions must presumably interact with any proteins involved in transposition.

The central region of most FB elements is made up of sequences related to the inverted repeats themselves[7]. FB4 is unusual in that its terminal repeats are separated by about 1,750 base pairs of unrelated DNA containing a potential gene coding for a 148 amino acid polypeptide. Most of this central region may itself be a transposable element since 33 nucleotides at its right-hand end occur as a near-perfect inverted repeat a short distance from the left-hand end (terminal inverted repeats are characteristic of all known prokaryotic and eukaryotic transposable elements).

The mechanism by which *copia*-like elements are able to redistribute themselves around the genome is by no means clear but the many properties they have in common with proviruses[3] suggest that it could involve circular intermediates[9] produced by a transcription–reverse transcription cycle. This is unlikely to be true of FB elements, however, and Potter (p.201) suggests that a transposase similar to those of bacterial transposons is involved and might been coded by the central region of FB4. Against this one has to remember that FB4 is the only FB element known to contain this sequence and its putative gene product might equally well be a transposase for the potential transposable element comprising this region.

The role (if any) of FB elements is unknown. Goldberg *et al.*[10] reported that an FB element lies at one end of one of the very large transposable elements (TE elements) described by Ising[8]. These are several hundred kilobases long and in some cases can be detected cytologically. The DNA involved can include respectable components of the genome, such as the white eye gene (*w*), unlike *copia*-like elements which seem to be genomic parasites. This raises the interesting possibility that any sequence bounded by two FB elements may potentially be transposable.

Inverted repeats can be detected in rapidly re-annealing DNA from many species and one may wonder whether these are similar to FB elements in the same way that endogenous proviruses seem to correspond to *copia*-like elements. In humans, the bulk of inverted repeat structures are comprised of two copies of the 300 base pair 'Alu' sequence in opposite orientation and separated by variable lengths of unrelated DNA[11]. Alu sequences make up a substantial component of human dispersed repetitive DNA and have counterparts in other species[12]. Jagadeeswaren *et al.*[13] have suggested that Alu sequences can themselves transpose but there is no indication that they behave like FB elements. It does not follow that humans and other vertebrates do not contain such elements, however, since even in *D. melanogaster* they make up less than two per cent of inverted repeats. All of this goes to show that there is still much work to be done in the study of repetitive DNA. □

1. Rubin, G.M. *et al. Cold Spring Harb. Symp. quant. Biol.* **45**, 619 (1981).
2. Tchurikov, N.A. *et al. Cold Spring Harb. Symp. quant. Biol.* **45**, 655 (1981).
3. Finnegan, D.J. *Nature* **292**, 800 (1981).
4. Schmidt, C.W., Manning, J.E. & Davidson, N. *Cell* **5**, 159 (1975).
5. Baker, R.F. & Thomas, C.A. in *DNA Insertion Elements, Plasmids and Episomes* (eds Bukhari, A.I., Shapiro, J.A. & Adhya, S.L.) (Cold Spring Harbor Laboratory, New York, 1977).
6. Potter, S., Truett, M., Phillips, M. & Maher, A. *Cell* **20**, 639 (1980).
7. Truett, N.A., Jones, R.S. & Potter, S.S. *Cell* **23**, 753 (1981).
8. Ising, G. & Black, K. *Cold Spring Harb. Symp. quant. Biol.* **45**, 327 (1981).
9. Flavell, A.J. & Ish-Horowicz, D. *Nature* **292**, 591 (1981).
10. Goldberg, M.L., Paro, R. & Gehring, W.U. *EMBO J.* **1**, 93 (1982).
11. Deininger, P.L. & Schmid, C.W. *J. molec. Biol.* **106**, 773 (1976).
12. Jelinek, W.R. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 1398 (1980).
13. Jagadeeswaren, P., Forget, B.G. & Weissman, S.M. *Cell* **26**, 141 (1981).

Seamounts and flexure of the lithosphere

from A.B. Watts

MORE than 75 years have passed since the first gravity measurements showed that the Hawaiian islands, in the Pacific Ocean, were associated with large positive gravity anomalies of up to a few hundred mGal. The American geodesists, Hayford and Bowie[1], showed in 1912 that these anomalies could be substantially reduced by taking into account isostasy. The form of isostasy they preferred was the Pratt model, which considered that the islands were locally compensated by lateral variations in crustal density. The model was popular with geodesists since it enabled them to readjust the triangulation system of the US: a task of immense practical importance in mapping and surveying. It was of comparatively little interest, however, to a leading geologist at the time, Barrell, because it did not account for the lateral strength of the crust. Putnam pointed this out to Hayford and Bowie who subsequently amended the Pratt model by extending the compensation from beneath the islands into surrounding regions. They showed that a regional model of isostasy could reduce the gravity anomalies over the Hawaiian islands further than could the Pratt model.

But it was not until 1926, when Vening Meinesz[2] made the first pendulum gravity measurements in a submarine, that it was possible to prove regional, rather than local, isostasy for the Hawaiian islands. The measurements showed that the positive anomalies over the islands were flanked by negative anomalies of up to 100 mGal, which Vening Meinesz first interpreted as a result of downbending or flexure of the crust due to the load of the Hawaiian volcanoes. He used a model of an elastic plate overlying a weak fluid substratum, similar to one developed by Hertz[3] to model the flexure of ice on ponds by skaters. Subsequently, Vening Meinesz demonstrated that a flexure model of isostasy reduced the gravity anomalies over the islands further than could either the Pratt model, favoured by Hayford and Bowie, or the Airy model, preferred by some European geodesists.

Since the development of the concept of plate tectonics the flexure model of isostasy has assumed a special significance. Plate tectonics views the outer layer of the Earth, or lithosphere, as consisting of several large plates which converge, move apart and slide past each other. The plates therefore correspond to the rigid layer argued for earlier, on isostatic grounds, by Barrell and Vening Meinesz. An important assumption of applying plate tectonics to the geological past is that the plates behave rigidly for long periods of time. Thus, one of the objectives of current flexure studies has been to determine the thickness of the rigid layer and whether it varies on long geological time scales.

The numerous oceanic islands and seamounts that occur in the world's oceans have proved particularly satisfactory loads for flexure studies. They are mainly of volcanic origin, form relatively quickly on the plates (as evidenced by the short times between eruptions) and occur in a variety of tectonic settings. Since WW II there have been a number of geological, geophysical and geodetic investigations of seamounts and oceanic islands so there is now a large amount of bathymetry, gravity, geoid and recent crustal movement data for them. By comparing these data with predictions based on elastic plate models it has been possible to estimate the flexural rigidity, and the equivalent elastic thickness of oceanic lithosphere, in different tectonic settings in the oceans[4–7].

Flexure studies at seamounts and

A.B. Watts is Professor of Geology in the Department of Geological Sciences, Columbia University, Palisades, New York 10964.

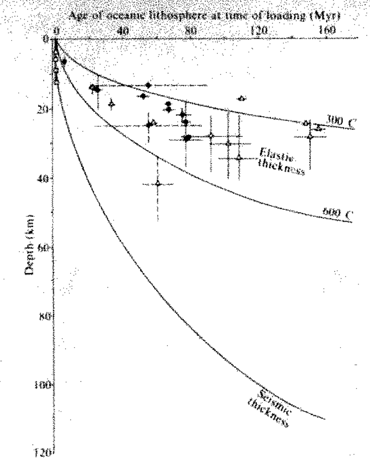
0028-0836/82/200182-02$01.00

oceanic islands have shown that the elastic thickness of oceanic lithosphere, which is significantly smaller than the seismic thickness, increases with the age of the lithosphere at the time of loading. The large range of ages represented by these features (1–56 Myr) suggest that on loading, the lithosphere rapidly relaxes from its initial short-term (seismic?) thickness to a long-term mechanical thickness. Subsequent relaxation has not been observed, but may occur on very long (> 56 Myr) time scales. The smallest values of elastic thickness (5–8 km) have been determined at features which form on young oceanic lithosphere, such as the western Walvis ridge guyots, while the largest values (17–37 km) have been determined at features which form on old oceanic lithosphere, such as the Hawaiian islands. These variations suggest[8] that as the oceanic lithosphere cools, it becomes progressively more rigid in its response to volcanic loads.

The flexure model, deduced from these studies, is referred to as 'elastic', even though the model parameters actually vary with time. The elastic thickness apparently changes rapidly following loading, approaching an asymptotic value that depends on the plate age. The lithosphere only appears elastic, therefore, after some time has elapsed ($\sim$1 Myr) and, even then, the thickness of the elastic layer is determined by the plate age.

The actual long-term mechanical properties of the lithosphere are, of course, likely to be more complex than would be predicted by an elastic model. A difficulty is that flexure studies only reveal information on the average mechanical properties of the lithosphere. The estimation of elastic thickness from gravity and geoid data, for example, is complicated by uncertainties in the volcanic load and infill densities, the crustal structure before flexure, and the lateral and vertical extents of lithospheric heating at the time of loading. But as Lambeck[9] and others have pointed out, gravity and geoid data are consistent with an elastic model in which the elastic thickness increases with age of the lithosphere at the time of loading. A more difficult problem, though, is that there is no maximum stress a purely elastic plate can store; yet the actual materials of the lithosphere are likely to have some ultimate strength. Goetze and Evans[10] have pointed out, however, that if a more realistic 'yield stress envelope' is used for the lithosphere (based on data from experimental rock mechanics), an elastic core would be expected to develop during flexure that corresponds closely to the elastic thickness determined using the elastic model.

Although an elastic model has been widely used in flexure studies[4–9], it has been criticized by some workers who prefer a viscoelastic (Maxwell) model. A viscoelastic model is characterized by an initial elastic response to a load, followed



Estimates of the elastic thickness (or mechanical thickness) of oceanic lithosphere plotted against age of the lithosphere at the time of loading. The solid symbols are estimates from studies by different workers at seamounts and oceanic islands, open symbols are estimates from studies at mid-ocean ridges (squares), river deltas (diamond) and deep sea trench–outer rises (triangles). As the age of most of the features range from a few Myr to several tens of Myr, flexure studies indicate that on loading the mechanically supportive part of the lithosphere relaxes from its short-term (seismic) thickness to a long-term mechanical thickness. The long-term thickness does not appear to change appreciably with time. Thus, flexure is likely to be an important controlling mechanism in the tectonic evolution of geological features in the oceans and continents.

by a viscous relaxation that rapidly increases with age of the load. Since the model incorporates a form of stress relaxation, the 'equivalent' elastic thickness would be expected to decrease as the load increases in age, the decrease being most apparent for long times (compared with the Maxwell relaxation time) and large loads. But elastic and viscoelastic models produce flexure curves of similar shape. Thus, the best way to distinguish between them is to determine whether the elastic thickness is controlled by the age of the load (viscoelastic model) or by the age of the plate (elastic model).

Walcott[4] first suggested, based on the results of flexure studies of continents and oceans, that the elastic thickness of the lithosphere depends on the age of the load, rather than on the plate age. A difficulty with his study, however, was that he did not satisfactorily account for differences in size or tectonic setting of the various geological features used in these studies. Recently, Lambeck[11] accounted for these differences in a study of the Society and Cook Islands in the South Pacific Ocean and argued for a dependence of the elastic thickness on age of these loads, that indicated a Maxwell time of about 10^7 yr for the oceanic lithosphere. The differences in age between the Society and Cook Islands are too small (1.4–3.1 Myr), however, to resolve satisfactorily whether there has been more relaxation beneath the older load (Cook) than the younger load (Society). The best example of a dependence of the elastic thickness on age of load occurs, in fact, along the Hawaiian–Emperor seamount chain; the younger Hawaiian ridge (4–18 Myr) shows the larger elastic thickness (37 km) and the older Emperor seamounts (44–58.5 Myr) the smaller thickness (22 km). But as Watts[8] pointed out, the small values for the Emperor seamounts can be more simply explained by an elastic, rather than a viscoelastic, model since these seamounts were originally emplaced on young, weak, oceanic lithosphere.

This discussion suggests, therefore, that a consensus is emerging from flexure studies at seamounts and oceanic islands favouring the elastic model as the most useful working hypothesis for the long-term mechanical properties of the lithosphere. The model can also explain (with some modification) the topography of the outer rise seaward of deep-sea trenches, the deep structure of large river deltas, the pattern of late glacial rebound due to the receding Laurentide ice caps and the patterns of gravity anomalies over orogenic belts and Precambrian granite plutons. The model is therefore currently enjoying a wide range of applications in geology that include the determination of the tectonic setting (on-ridge or off-ridge) of intraplate volcanoes, the nature of the control on stratigraphical sequences in sedimentary basins and the gravity and geoid effect of deep processes beneath the lithosphere, such as layered mantle convection.

Clearly, the elastic model remains a working hypothesis that needs to be tested in the future. A critical test would be to determine the actual seismic thickness of the crust and lithosphere in the region of a large volcanic load, such as the Hawaiian islands. Recently, two-ship multichannel seismic experiments, using long arrays and repetitive sound sources, have been carried out in the western and eastern Pacific Ocean[12] that have revealed nearly continuous reflections from the 'Moho' discontinuity at the base of the oceanic crust. Similar experiments in the central Pacific, near the Hawaiian islands, offer the most promise, therefore, to test the flexure model of isostasy during the next decade. □

1. Hayford, J.F. & Bowie, W. *Spec. Publ.* **10** (Coast and Geodetic Survey, Washington DC, 1912).
2. Vening Meinesz, F.A. *Publ. Neth. Geodetic commun.* (Delftsche Uitgevers Maarschappij, Delft, 1948).
3. Hertz, H. *Wiedemann's Annalen* **22**, 449 (1884).
4. Walcott, R.I. *J. geophys. Res.* **75**, 3941 (1970).
5. Watts, A.B. & Cochran, J.R. *Geophys. J.* **38**, 119 (1974).
6. McNutt, M. & Menard, H.W. *J. geophys. Res.* **83**, 1206 (1978).
7. Roufosse, M.C. *Proc. 9th GEOP. Conf. Dept Geodetic Sci. Rep.* **280**, 261 (1978).
8. Watts, A.B. *J. geophys. Res.* **83**, 5989 (1978).
9. Cazenave, A., Lago, B., Dominti, K. & Lambeck, K. *Geophys. J.* **63**, 233 (1980).
10. Goetze, C. & Evans, B. *Geophys. J.* **59**, 463 (1979).
11. Lambeck, K. *Earth planet. Sci. Lett.* **55**, 482 (1981).
12. Stoffa, P.L., Buhl, P., Herron, T.J. & Ludwig, W.J. *Mar. Geol.* **35**, 83 (1980).

Gene rearrangement and the generation of diversity

from Miranda Robertson

THE eukaryotic genome has rececently begun to seem less stable, both in evolution[1] and in ontogeny[2], than many biologists had assumed. These trends prompted the Institute of Genetics to invite to its annual Spring meeting speakers on six genetically labile systems: the MHC locus and the immunoglobulin genes of mammals, the genes encoding the surface antigens of trypanosomes, and the genes of plant and animal tumour cells. The MHC locus, which illustrates the evolutionary lability of the vertebrate genome, will be discussed in detail in a later article. What follows here is concerned with the generation of diversity on a much shorter time-scale, in antibodies during the lifetime of vertebrates, and in the surface antigens of parasitic trypanosomes in the course of an infection.*

Both the antibodies of vertebrates and the surface antigens of trypanosomes have evolved the means of extremely rapid diversification through DNA rearrangements leading directly to changes in gene expression. But the selective pressures on antibody-producing vertebrate lymphocytes are different in fundamental ways from those acting on trypanosomes; and those differences may be reflected in the mechanisms by which the genetic changes are achieved. Speakers at Cologne explored both the causes and the consequences of variation in these two systems — and touched on some implications for the normal and abnormal development of multicellular animals.

Deletion and recombination in immunoglobulin

The somatic rearrangement of the DNA of antibody genes takes place in lymphocytes of the lineage whose end-product is the differentiated antibody-secreting B cell, or plasma cell. Its effect is to join the two normally separate gene segments (the V and J segments) to encode the variable region of the light chains of the antibody molecule, and of three separate gene segments (the V,D and J segments) to encode the variable region of the heavy chain (see summary diagram in ref.2). It is no longer disputed that the variability of the variable regions of antibodies arises from the largely random assembly of each of the coding sequences from a large pool of germ-line V segments, a smaller pool of J segments, and (for heavy-chain V regions) a pool of D segments whose extent is still unknown. There is however vigorous debate about the mechanisms by which the recombination takes place.

The only obvious clue yielded by the recombining DNA itself is the presence of conserved 'signal' sequences flanking each of the separate gene segments (see legend to Figure 1). Because these sequences contain inverted DNA repeats, Leder[3] and Tonegawa[4], both in 1979 working on the light-chain genes of mice, proposed that the V-J join results from the formation of a stem structure through intrachromatid base-pairing in which the DNA intervening between the V and the J loops out and is deleted (Fig. 1*a*). In support of this proposal, they reported the absence of the intervening DNA in immunoglobulin-secreting myeloma cells.

The deletion theory however very soon encountered a challenge in the work of Steinmetz *et al.*[5], who found in a mouse myeloma a fragment of the deleted DNA containing a recognizeable pair of back-to-back signal heptanucleotides. It is clear from Fig. 1*a* that such a configuration could be generated by the excision of the Leder/Tonegawa loop-and-stem if it were ligated at its base to form a circle. But the persistence of such circles in antibody-secreting cells is not consistent with the disappearance of the intervening DNA from the cells analyzed by Leder and Tonegawa.

Furthermore, the persistence of back-to-back signal sequences is quite general: they have now been found in many more myelomas and hybridomas, both by Walfield *et al.*[6] and by Van Ness and colleagues (M. Weigert, Institute of Cancer Research, Philadelphia). It has thus become necessary to think again about the mechanism of V-J joining; and one alternative, advocated at Cologne by Weigert and by J. Höchtl (from the laboratory of H. Zachau at the University of Munich where the first back-to-back signal fragment was found) is mitotic recombination resulting in the unequal exchange of DNA between sister chromatids (Fig. 1*b*).

The essential difference between their theory and the Leder/Tonegawa theory is that while the loop-and-stem mechanism

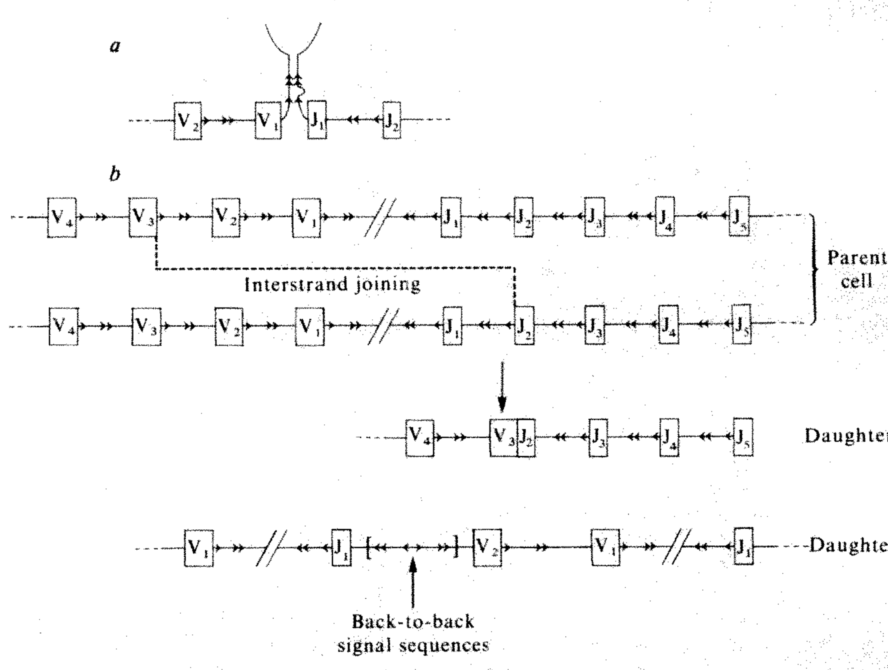


Fig 1. The joining signals for the V and J segments of light-chain genes are a conserved heptanucleotide (>) followed by a conserved nonanucleotide (»), separated by about twelve base-pairs of non-conserved DNA in the case of the V-segment signal, and about 24 base-pairs in the J-segment signal.

For the heavy chain, which is not illustrated here, the conserved sequences are the same but both V and J signals have a 24-base-pair non-conserved spacer. The D segments, which in the assembled coding region come between them, are flanked on both sides by signals with a 12-base-pair spacer. This arrangement has led to the enunciation of the so-called 12-24 rule for segment joining: joining is held to occur only between signals with different spacer lengths.

(*a*) Hypothetical loop-and-stem structure giving rise to a V-J join. If the loop were ligated at the base of the stem, the result would be a back-to-back pair of signal heptanucleotides.

(*b*) The consequences of mitotic recombination with an unequal exchange of material between sister chromatids. One daughter inherits a chromosome with a functional V-J join, the other a back-to-back heptanucleotide pair and the potential for a future V-J join. Only one pair of sister chromatids is shown for the dividing parent, and only one chromosome of each daughter: diploid cells will, of course, have two.

Miranda Robertson is associate editor of Nature.

---

*The Spring Meeting on Molecular Biology, whose theme was the Genetic Basis of Diversity, was held in Cologne on Feburary 24-27, 1982.

0028-0836/82/200184-03$01.00

entails the redistribution of DNA within a chromatid, interstrand joining between sister chromtids entails redistribution between chromatids, with subsequent segregation of the two chromatids to different daughter cells.

Höchtl and Weigert presented circumstantial evidence for such segregation at Cologne. Höchtl *et al.*[7] have now sequenced back-to-back signal fragments from three myelomas, and they find that the flanking regions of the back-to-back signals are not those of the V and J segments actually rearranged int he myeloma cells. This can easily be explained if the myelomas had inherited a chromosome such as that illustrated for daughter 2 in Fig. 1*b*, productive rearrangement of the expressed segments occurring either by a second recombination involving the same chromosome, or by a rearrangement on the other chromosome.

Mitotic recombination can also effortlessly explain the 'deletion' of DNA seen by Leder and Tonegawa, and by Weigert in some cell lines, since after two or more rounds of mitotic recombination a proportion of cells will inherit two shortened chromosomes. Equally, however, it follows that the total DNA content of a population of normal B lymphocytes should be conserved. This prediction has been tested by Van Ness *et al.*[8] who have established that most of the DNA upstream of the J region is retained in a population of normal B lymphocytes from mouse spleen.

Multiple rearrangements

But there is one feature of the data on V-J joining whose explanation is not so obvious. That is the preponderance of rearrangements involving J_1. This is reflected both in the proteins produced by myelomas and in the sequenced back-to-back fragments, all of which so far prove to contain flanking DNA from the J_1 segment. Weigert suggests that this may be a consequence of rearrangements of upstream J segments remaining on the chromosome containing the back-to-back signal fragment: the chromosome of daughter 2 in Fig. 1*b* illustrates such an upstream J. The downstream J segments remaining on daughter 1 would never be rearranged because rearrangement ceases once a V segment has been productively joined to a J. This would lead to an overall bias in favour of J_1.

On the other hand, if multiple rearrangements are frequent then the back-to-back signal fragments would no longer be expected to correspond in a one-to-one fashion with the productively rearranged segments in a given cell, and one of the stronger arguments for exchange between sister chromatids would be weakened. Indeed Baltimore, who has evidence that immunoglobulin genes are capable of multiple rearrangements, places quite a different interpretation on the back-to-back signal sequences that he, too, finds associated with rearranged light-chain genes. His evidence for multiple rearrangements comes, not from light-chain genes but from a recent analysis of heavy-chain V-D-J joining by Alt and Baltimore[9].

By following the fate of the D segments in a pre-B cell line transformed by murine leukaemia virus (MuLV) and still undergoing DNA rearrangement, he and his colleagues have been able to show that multiple rearrangements of D segments are possible; in one case three separate joining events between D and J segments alone must have occurred on a single chromosome. The final result of these rearrangements is a D segment sandwiched between two J segments and therefore incapable of participating in heavy-chain production. But two features of the rearranged chromosome have important general implications for the mechanism of immunoglobulin-gene rearrangement. First, a pair of back-to-back heptamers has been generated upstream from the joined segments; and second, the DNA between the heptamers and the downstream margin of the D segment is inverted (Fig. 2). While the heptamers could in principle be explained by recombination between sister chromatids, the inversion cannot, since it would entail breaking the chromosome. Baltimore believes that the back-to-back heptamers in rearranged light-chain DNA, too, may be generated by the inversion of V segments during V-J joining. (The orientation of the V segments relative to the J segments in germ-line DNA is not in fact known.) At this stage, however, no mechanism for V-J joining can be definitively ruled out.

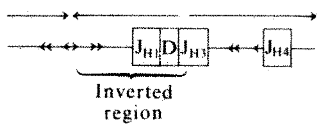


Fig 2. Abortively rearranged heavy-chain DNA of a virus-transformed cell line described by Alt and Baltimore[9]. the arrows indicate the relative orientation of the DNA.

More diversity by mutation and conversion

In any case there is more diversity in the variable-region coding sequences than either mitotic recombination or intrachromosomal deletion can explain on its own. It turns out that somatic mutation occurs perhaps during and certainly after the DNA rearrangements that join the different gene segments[2,10]; this can result in a very large number of substitutions — as many as 44 in a single V region. L. Hood (California Institute of Technology), on the basis of a detailed analysis of expressed and germ-line sequences encoding the V regions of antibodies that bind the hapten phosphorylcholine (PC), has concluded that there must be a special mechanism for generating localized mutations, since almost all of them are confined to the DNA encoding the V region and its immediate vicinity[11]. He calls the phenomenon hypermutation.

It can, however, be difficult to be certain that the substitutions are due to mutation and not recombination between members of the extensive V-gene family. Hood's evidence rests on the identification of a family of germ-line V segments that hybridizes strongly with the expressed anti-PC V sequences and not with other germ-line genes; and furthermore he finds substitutions in the J region, for which all the germ-line sequences are known.

The danger of postulating mutation in the absence of complete information on germ-line sequences was illustrated by collaborators of K. Rajewsky and K. Bayreuther (University of Cologne), who in similar studies have discovered a V-region variant arising *in vitro* in which what was apparently a cluster of mutations in the heavy-chain V region proved to be identical to a part of the germ-line sequence of a different V segment. They have interpreted this striking discovery as a possible instance of gene conversion — a mode of recombination increasingly invoked to account for otherwise inexplicable blocks of homology.

The question of whether conversion makes a significant contribution to antibody diversity *in vivo* however remains open.

Structure and function of antibodies

The random recombination and subsequent mutation, or conversion, of immunoglobulin variable-region gene segments guarantee diversity of structure, but they by no means guarantee diversity of function. The extremely ill understood relationship between the two has been the chief preoccupation of S. Rudikoff (National Institutes of Health) who reported on the basis of an analysis of variants of anti-PC antibodies, that amino-acid substitutions may make remarkably little radical difference to antigen-binding specifity. Hood hazarded that the choice of germ-line segments probably broadly determines the character of the antigen bound, while mutations (or conversions) 'fine-tune' the binding properties of the antibody.

An extraordinarily elegant illustration of this principle was presented by W. Gerhard (Wistar Institute, Philadelphia), in whose laboratory a large panel of hybridomas has been constructed from individual inbred mice immunized with influenza A virus, and the monoclonal antibodies tested for binding to a range of mutant viral haemagglutinins. From among these antibodies, one group could be identified, on the basis of restriction mapping, as a family probably derived from the same re-arrangement of germ-line segments, and thus presumably from the same precursor B cell. The antigen-binding properties of most of them could, however, be clearly distinguished by single-point mutant

haemagglutinins, each of which was recognized by only two or three of the antibodies.

What this means, in practical terms, is that somatic mutation in memory B cells should enable an immunized mammal to keep pace with the continous minor antigenic drift of the influenza haemagglutinin molecule. It may also explain why, by contrast, the more dramatic antigenic shift that occurs every few years can give rise to pandemics[12]: the more profound structural effects of the shift may demand a fresh combination of germ-line segments and leave the animal temporarily unprotected.

Recombination and expression of genes for trypanosomes surface antiges

Like the influenza virus, the African trypanosome has evolved the ability to change its antigenic coat proteins and thus evade the immune armoury of its vertebrate host. Indeed, the surface antigens of trypanosomes change sequentially in the course of a single infection and generally outflank the immune system altogether. But the parallel between trypanosomes and influenza viruses is in many ways less striking than the parallel between trypanosomes and the B lymphocytes of their victims. Trypanosomes, like B lymphocytes, have a family of related genes encoding surface proteins; and, like B lymphocytes, they can vary the expression of the genes by rearrangements of the chromosome containing them[13].

The manner of the rearrangement may however be less like that of antibody genes than that of the mating-type genes in yeast, since the expression of a variable surface glycoprotein (VSG) gene in trypanosomes seems to be preceded, at least in some cases, by duplication and transposition[13], whereas no duplication (so far as is known) accompanies the rearrangement of antibody V segments. P. Borst (University of Amsterdam) has suggested that the duplication-transposition of trypanosome VSG genes could be another instance of gene conversion, in which the gene in the expressed locus — the so-called expression-linked copy — is converted sequentially by the other, co-called basic copy, genes. (The conversion event posulated by the Cologne group in immunoglobulin V segments occurs after gene rearrangement and not as part of its mechanism.) There is moreover no particular reason to expect that the same recombination mechanism will operate in B lymphocytes, which do not need to maintain their genomes intact for transmission to the next generation, and trypanosomes, which do.

Transposition and activation of genes

A significant point of similarity however is that in both cases the rearrangment of DNA results in the activation of the rearranged gene. The question is, how? In the case of antibody genes, it is known that there are promotors upstream of not only the unrearranged constant-region segments at the transcribed locus but each of the germ-line V segments as well. The constant-region sequences are transcribed at low levels from unrearranged chromosomes in both B and T lymphocytes, suggesting that the locus is constitutively active in the lineage. The unrearranged V segments by contrast are silent. Van Ness *et al.* have established that recombination of the V segment (and its promotor) with the J and C segments increases the transcriptional activity of the C region severalfold[13]. They conclude that the total transcriptional activity may depend on the combined effects of both V segment promotors, and the structure of the chromatin at the site of the C segment — the appropriate chromatin structure being the necessary condition for expression. (There is in fact evidence for a region of DNase 1 sensitivity upstream of the C-region genes[15].)

Similar factors may determine VSG gene expression in trypanosomes, though Borst believes the powerful promotor may be at the expression site rather than at the basic copy site. He reported at Cologne that the mRNA transcribed from the expressed gene contains a 5′ sequence that does not correspond to the 5′ flank of the basic-copy gene, and suggested that it is derived from the DNA at the expressed locus, which may contain an upstream promotor[16]. This hypothesis could clearly be tested by cloning the DNA of the expression-linked copy.

But trypanosomes in have in any case more than one way of changing the expression of their VSG genes. Williams *et al.*[17] have found rearrangement of the expressed gene without evidence of duplication; and S. Longacre (Pasteur Institute, Paris) reported that duplicated genes in *T. equiperdum* may persist unexpressed while a second duplicated gene in a different locus is transcribed.

If the rearrangement of trypanosome VSG genes does not bring them under the control of a promotor, perhaps, like antibody genes, they bring their own promotors with them and activation is due to the state of the surrounding chromatin.

This is the mechanism suggested by Nasmyth *et al.*[18] to account for the alternating expression of the a and α mating-type genes of yeast. Each yeast cell carries one silent copy of each of the two related genes, a duplicated a or α gene being expressed in alternate generations at the MAT locus. There is some evidence that the chromatin at the MAT locus is more loosely packed than at the silent loci.

But the importance of chromatin structure may not be confined to its effect on transcriptional activity. Mating-type genes, VSG genes, and immunoglobulin V segments are not transposed to random sites in the genome, and the state of the chromatin may help to determine where they recombine. In every case, new sequences are transposed to an already active site, and Van Ness *et al.* have suggested that, in the case of antibody genes, transcriptional activity may be a prerequisite for recombination. In this context, it may be extremely significant that in mutant yeast strains in which the normally silent mating-type loci are transcribed, these loci, like the normally active site, switch between a and α sequences; in wild-type strains, these silent loci are never expressed and do not switch[19].

Transformation and differentiation

Are these variations in position and expression of related genes a peculiarity of a handful of highly specialized genetic systems, or have they (as Strathern and Herskowitz have suggested for yeast[20]) important implications for the study of differentiation in general? It has already been pointed out[21] that the existence of three families of mobile promotors on the chromosomes containing the three pools of V segments poses a serious oncogenic threat. Since it is now generally held that cancer may be caused by normal cellular genes promoted to abnormally high levels of activity, the arrival of a V segment complete with promotor in the vicinity of such a gene might lead to cancer; and indeed 'here is evidence that some leukaemias may arise in just this way[21].

Speculations on the part played by labile gene families in normal development however are so far largely unconstrained by evidence. Thus unfettered, Hood launched a pan-explanatory schema in which embryogenesis is guided by varying surface markers differentially and sequentially expressed on cells with different developmental destinies, each additional surface marker further narrowing the differentiation pathway open to its bearer — just as differential expression of antibody V-regions narrows the interactions open to a lymphocyte. Such is the present pace of progress that today's imaginative leap may land in tomorrow's text-book. □

1. see Jeffreys, A.J. & Harris, S. *Nature* **296**, 9 (1982).
2. Robertson, M. *Nature* **290**, 625 (1981).
3. Max, E.E., Seidman, J.C. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3540 (1979).
4. Sakano, H., Hüppi, K., Heinrich, G. & Tonegawa, S. *Nature* **290**, 288 (1979).
5. Steinmetz, M., Altenburger, W. & Zachau, H.G. *Nucleic Acids Res.* **8**, 1709 (1980).
6. Walfield, A., Selsing, E., Arp, B. & Storb, U. *Nucleic Acids Res.* **9**, 1101 (1981).
7. Höchtl, H.J., Müller, C.R. & Zachau, H.G. *Proc. natn. Acad. Sci. U.S.A.* **79**, 1383 (1982).
8. Van Ness, B.G., Coleclough, C., Perry, R.P. & Weigert, M. *Proc. natn. Acad. Sci. U.S.A.* (1982).
9. Alt, F. & Baltimore, D. *Proc. natn. Acad. Sci. U.S.A.*, in the press.
10. Alt, F. *et al. Cell* **27**, 381 (1981).
11. Kim, S. Davis, M., Potter, P. & Hood, L. *Cell* **27**, 573 (1981).
12. Webster, R.G., Laver, W.G., Air, G.M. & Schild, G.C. *Nature* **296**, 115 (1982).
13. Hoejmakers, J.H.J. *et al. Nature* **284**, 78 (1980).
14. Van Ness, B.G. *et al. Cell* **27**, 593 (1981).
15. Storb, U., Arp, B. & Wilson, R. *Nature* **294**, 90 (1981).
16. Bernard, A. *et al. Cell* **27**, 497 (1981).
17. Williams, R.O., Young, J.R. & Majwa, P.A.O. *Nature* **282**, 847 (1979).
18. Nasmyth, K. *et al. Nature* **289**, 244 (1981).
19. Klar, A.J.S., Strathern, J.N. & Hicks, J.B. *Cell* **25**, 517 (1981).
20. Strathern, J.N. & Herskowitz, I. *Cell* **17**, 371 (1979).
21. Klein, G. *Nature* **294**, 313 (1981).

REVIEW ARTICLE

Chemistry of hot springs on the East Pacific Rise and their effluent dispersal

J. M. Edmond*, K. L. Von Damm, R. E. McDuff & C. I. Measures

Department of Earth and Planetary Sciences, Massachusetts Institute of Technology, E34-201, Cambridge, Massachusetts 02139, USA

Cold groundwater circulating through submarine ridge crests reacts with hot basalts at temperatures in excess of 300 °C. The resulting solutions ascend and emerge as hot springs on the sea floor. This ridge crest hydrothermal activity is pervasive and chemically uniform. The fluxes of the various elements injected into, or removed from, the ocean are substantial compared with fluvial transport. The dispersal of the effluent is shown to be controlled by global circulation at mid-depth.

LARGE-SCALE, high temperature, hydrothermal activity is a ubiquitous concomitant to the production of oceanic crust at submarine ridge crests. The establishment of this fact has been one of the most important recent developments in oceanography and, indeed, in the earth sciences as a whole. Molten rock is injected at the centres of sea floor spreading at temperatures of around 1,200 °C. The pre-existing crust is highly permeable because of faulting and of fissuring induced by thermal contraction. Cold 'groundwater' derived from the overlying ocean can thus circulate through the crust and penetrates the zone of active intrusion along cracking fronts[1]. Here reactions take place between the seawater and the hot basalt at temperatures in excess of 300 °C. The resulting solutions ascend to the sea floor where they emerge as hot springs.

Geological evidence

The phenomenon was first proposed by Elder[2] from analogy with geothermal systems associated with continental igneous activity. At about the same time Bostrom and Peterson[3] invoked ridge crest hot springs of undefined scale as the source of the elevated concentrations of iron and manganese which they had shown to be a characteristic feature of the sediments on the crestal zone of the East Pacific Rise. Such a local source was required by their observations since abyssal sediments remote from the axes showed no such enrichments. As data were accumulated on the chemistry and mineralogy of basalts dredged from the ridge crests it became clear that, in many cases, they had been pervasively altered by reaction with seawater[4,5]. The partitioning of oxygen isotopes between coexisting secondary minerals, the reaction products of this alteration, demonstrated that the temperatures of the interactions ranged up to several hundred degrees centigrade[6].

Parallel investigations in ophiolites, slabs of oceanic crust tectonically emplaced onto the continents, gave similar results and established that seawater must penetrate over 5 km into the intrusion zone at the ridge crests[7]. Metalliferous sediments were found to be generally associated with the uppermost pillow lavas in the ophiolite sections and to be of hydrothermal origin[8,9]. Ore deposits composed of massive accumulations of pyrite (FeS_2) in the same terrains were shown to require very large volumes of high temperature solutions for their transport and emplacement[10].

Results from the Deep Sea Drilling Project demonstrated that metalliferous sediments are the usual basal deposit in the contemporary oceanic sediment column[11]. The underlying basalts, where sampled, showed evidence of reaction with seawater over a wide range of temperatures[12]. Detection of the primordial, non-radiogenic, isotope of helium, ^3He, at mid-depth in the Pacific water column[13] provided a strong additional indication that hydrothermal activity is occurring along the crest of the East Pacific Rise.

* Present address: Institut de Physique du Globe, Université de Paris VI, France.

Quantitative estimates of the global scale of this seawater convection began to emerge in the mid-1970s from comparison of the measured conductive heat loss from young oceanic crust with that predicted from the thermal models of the plates[14]. A pronounced negative anomaly was observed everywhere[15]. The total deficit for all the active ridges is estimated at between 2 and 8×10^{19} cal yr^{-1} (ref. 16). Assuming that this heat was removed by convective flow at typical temperatures of 300 °C, as suggested by the ^{18}O data[6], then a volume of water equivalent to the whole ocean must circulate through the high temperature intrusion zone in the ridge axes (that is through the 300 °C isotherm) every 8–10 Myr.

The chemical importance of this phenomenon was illustrated in the earliest experimental work on seawater–basalt interactions at elevated temperatures and sea floor pressures[17]. Drastic changes were observed in the compositions of the solutions during the reactions. Magnesium and sulphate were rapidly depleted and potassium, calcium and silica enriched. Simple calculations indicated that ridge crest hydrothermal activity might be a major, hitherto unconsidered, component of the chemical mass balance of the oceans over geological time[15].

Chemistry of ridge crest hot springs

Finding and sampling actual examples of ridge crest hot springs proved to be complex. Although numerous indications of hydrothermal activity had been detected in the water column over the axial zone[18,19], it was not until 1977 that a comprehensive exploration programme at the Galapagos Spreading Centre in 86° W in the eastern equatorial Pacific succeeded in locating and collecting water from discrete vents[20]. This required use of high resolution bathymetric charts, an accurately navigated, large-field, survey camera and the research submersible *Alvin*.

The hot springs at the Galapagos Spreading Centre were found to occur in extensive fields[21] associated with diverse and exotic fauna[22]. The exit temperatures were low, in the range 6–20 °C above that of the ambient water column at the axial depth (2.05 °C, 2,500 m). However, systematic investigation of the chemistry of the solutions demonstrated that these low and variable temperatures were actually an artefact of large-scale, sub-surface mixing between a very high temperature component and 'groundwater' indistinguishable in properties from the ambient seawater[23,24]. Individual chemical elements in solution respond differently to this process. The hydrothermal solutions were cool, contained variable amounts of H_2S (10–160 μmol kg^{-1}) and were mildly acidic[25]. Elements which do not form insoluble phases in such conditions are unaffected by the mixing process.

The experimental data[17] indicated the complete removal of dissolved magnesium during high temperature seawater–basalt reactions. Assuming that the fluids sampled at Galapagos represented a simple mixture between cold 'groundwater' and a single hydrothermal endmember then the temperature of the latter could be estimated by extrapolating the observed negative

0028-0836/82/200187–05$01.00

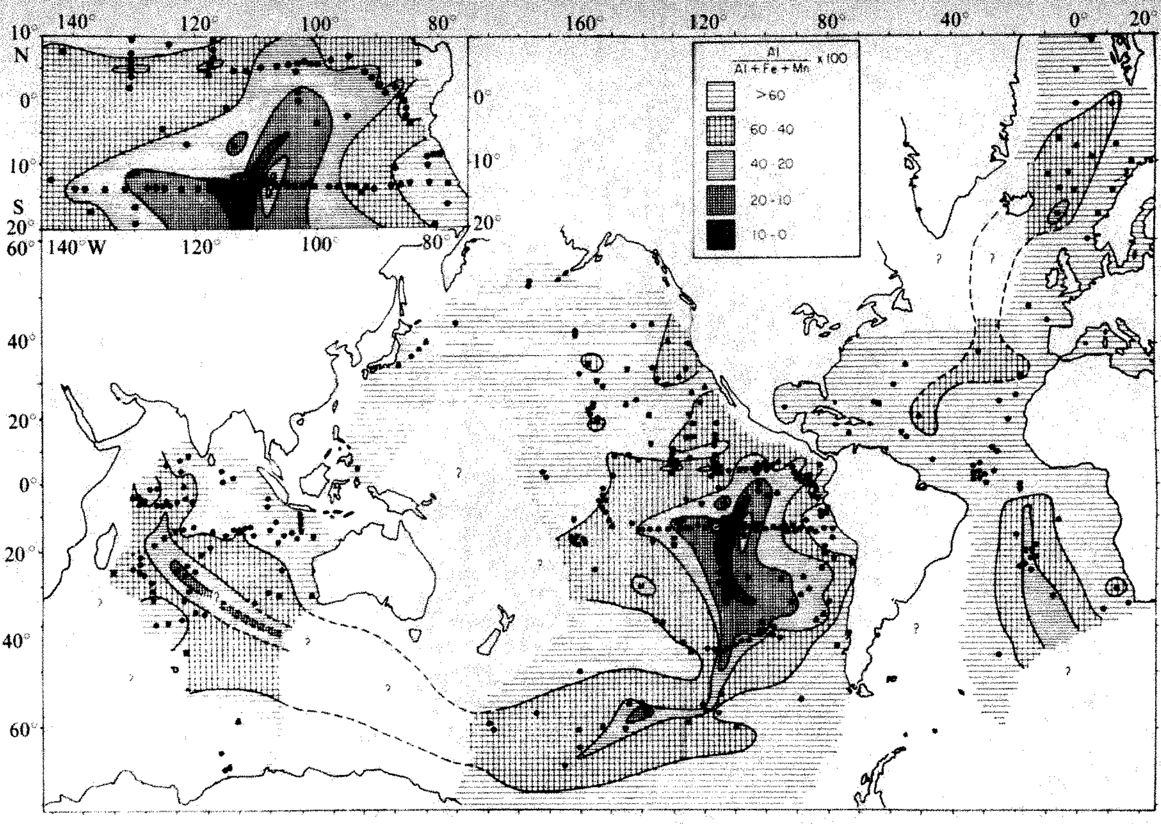


Fig. 1 Distribution of metalliferous surface sediments in the ocean as indicated by the ratio Al/(Al+Fe+Mn). As these deposits are poor in aluminium relative to continental detritus their abundance is inversely proportional to the value of the ratio. Figure reproduced from ref. 38.

magnesium–temperature relationship in the samples to zero concentration. An intercept was obtained at ~350 °C (ref. 23). This was confirmed by application of the quartz solubility criterion—quartz geothermometry[26]—to the parallel data for dissolved silica. Extrapolation of the concentration–temperature trends for the other conservative species then allowed estimates to be made of the composition of this putative 350 °C endmember (Table 1). Elements which form insoluble precipitates—sulphides and oxides—in the conditions observed in the Galapagos solutions showed sharp depletions with increasing temperature, their concentration levels extrapolating to zero between 30 and 40 °C (ref. 24). Not only were these elements—Cu, Ni, Cd, Se, U, Cr—being removed from the high temperature component of the fluids but also from the admixed groundwaters. It was not possible, therefore, to obtain any information on the concentrations of these potentially ore-forming elements in the high temperature solution.

Despite such complexities some remarkable results were obtained from the Galapagos data: these supported far-reaching conclusions as to the global significance of ridge crest hydrothermal activity. Most important are the fluxes involved. For ^{3}He the ratio to transported heat is 2.2×10^{-17} mol cal^{-1} (ref. 27). Integration of the water column ^{3}He anomaly, coupled with ^{14}C-calibrated models for the oceanic overturn rate, results in a flux of ^{3}He into the ocean, necessary to maintain the anomaly, of 1,100 mol yr^{-1} (ref. 28). This is in good agreement with the terrestrial budget which is dominated by a loss term due to gravitational escape to space. If it is assumed that all the ^{3}He delivered to the oceans is from hot springs in the ratio to heat observed at Galapagos then the associated heat flux can be calculated. This independent estimate agrees exactly with the geophysical results[27].

The apparent success of this approach coupled with appreciation of the close constancy in composition of the hydrothermal reactants, tholeiite basalts and seawater, justified similar calculations for the other dissolved species. The resulting values for elemental fluxes into and out of the axial zone are in many cases comparable with or greater than those estimated for fluvial transport of the products of continental weathering[23]. Thus the ridge is the major sink for magnesium and sulphate, the major source for lithium, rubidium and manganese and a substantial contributor of calcium, barium and silica. Several long-standing problems in the construction of oceanic chemical balances are resolved by inclusion of the hydrothermal term. The situation for the non-conservative elements cannot be evaluated[22] as the global distribution of exit temperatures—the average extent of sub-surface mixing—is, of course, unknown.

The validity of the deductions based on the Galapagos data was confirmed by the discovery, 2 years later, of hydrothermal fields on the crest of the East Pacific Rise in 21° N, south of the mouth of the Gulf of California[29,30]. Here, in depths of 2,500 m, solutions at temperatures of 350 °C were observed issuing from the sea floor. They had associated with them constructional features, chimneys up to 10 m high, composed

Table 1 Comparison of the estimated composition of a 350 °C hydrothermal endmember based on extrapolation of the Galapagos data with that observed at 21° N

	Galapagos	21° N	Seawater
Li (μmol kg^{-1})	1,142–689	820	28
K (mmol kg^{-1})	18.8	25.0	10.1
Rb (μmol kg^{-1})	20.3–13.4	26.0	1.32
Mg (mmol kg^{-1})	0	0	52.7
Ca (mmol kg^{-1})	40.2–24.6	21.5	10.3
Sr (μmol kg^{-1})	87	90	87
Ba (μmol kg^{-1})	42.6–17.2	95–35	0.145
Mn (μmol kg^{-1})	1,140–360	610	0.002
Fe (μmol kg^{-1})	+	1,800	—
Si (mmol kg^{-1})	21.9	21.5	0.160
SO_4 (mmol kg^{-1})	0	0	28.6
H_2S (mmol kg^{-1})	+	6.5	0

The ranges in the Galapagos results derive from the different trends for composition versus heat observed between individual vent fields[23]. +, Non-conservative to sub-surface mixing; —, seawater concentration not accurately known.

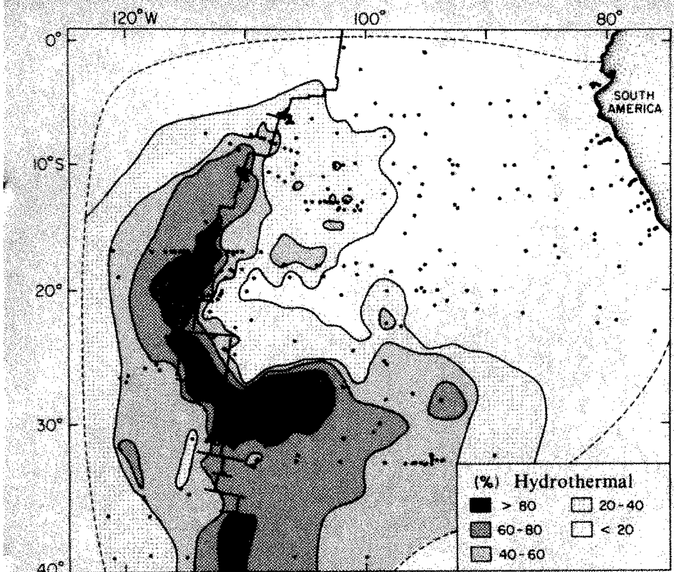


Fig. 2 The weight percentage of the hydrothermal component on a carbonate and salt-free basis in surface sediments of the Nazca Plate[33]. The abundance of this component is calculated from chemical data for the bulk sediment using a normative composition model.

predominantly of iron, zinc and copper sulphides. These grow from the hydrothermal solutions as they mix with the ambient seawater. However, most of the precipitation occurs as very fine-grained particles entrained in a turbulent plume rising tens of metres above the orifices of the springs[30].

Sampling of these hot springs established that they are indeed the compositional endmember postulated on the basis of the Galapagos data[31]. The waters exit as clear, homogeneous solutions within a few degrees of 350 °C. They are completely depleted in magnesium and sulphate and appear to be saturated with quartz. The concentrations of the conservative species are close to the Galapagos extrapolations (Table 1). The ^{3}He/heat ratio is identical to that from Galapagos[32]. The validity of the flux calculations is substantially confirmed.

These undiluted solutions also preserve the signatures of elements precipitated during Galapagos-type sub-surface mixing. The iron concentration is 1.8 mmol kg^{-1}, the sulphide 6.5 mmol kg^{-1}. The Fe/Mn ratio is close to 3, coincidentally identical to the average in ridge crest metalliferous sediments[33]. This low value, as compared with that in basalt (50–60), indicates the importance of iron oxidation in the reduction of seawater sulphate in the reaction zone. There is also a possible solubility control by pyrrhotite (FeS) although existing thermodynamic data are inadequate to be certain of this[34].

The isotopic data as well as confirming the ^{3}He/heat ratio[30], show that the substantial enrichment of CO_2 in the solutions is of mantle origin ($\delta^{13}C = -6‰$). The ^{18}O in the existing fluids[32], +1.6‰ reflects basalt equilibrium, confirming the previous results for dredged rocks and ophiolites. The isotopic composition of the dissolved strontium (ref. 35 and R.E.McD. and J.M.E., unpublished data) is essentially the same as in the basalts, 0.703. It has been realized for some time[36,37] that a large flux of non-radiogenic strontium is required to balance the river input (>0.712) and maintain the oceanic value at 0.709. The ridge component is sufficient to close this balance, resolving the problem and supporting the generality of the flux calculations.

Dispersal of hydrothermal effluent

The hot springs at 21° N are ore-transporting solutions such as are responsible for the massive sulphide deposits found in ophiolites and greenstone belts[10]. However, the abundance of springs at 21° N is apparently insufficient to produce more than small accumulations[30,31]. In addition their position in the local topography and hence in the ambient current field prevents significant local deposition of sedimentary sulphides from the plumes. The particulate sulphides and the large amounts of dissolved, divalent manganese are therefore dissipated by the general mid-depth circulation, and on oxidation in the water column precipitate out regionally contributing to the metalliferous sediments characteristic of the East Pacific Rise. During their chemical evolution in the plumes, the water column and sediments these Fe–Mn oxides must interact extensively with other elements in solution producing the diverse compositional and isotopic signatures observed[11].

While the chemistry involved in this evolution is not understood beyond the descriptive level, a significant body of information does exist on the distribution of these hydrothermal sediments in the Pacific. The original maps of Bostrom *et al.*[38] showed a pronounced asymmetry in the index Al/(Al+Fe+Mn) about the axis of the East Pacific Rise (Fig. 1). The magnitude of this ratio is inversely proportional to the abundance of the hydrothermal component which is, like the hot spring solutions and the water column, depleted in aluminium. More detailed work by Dymond[33] on the surface sediments of the south-east Pacific (Nazca Plate) confirms the southeastern lobe in Bostrom's distribution, centred on 30° S (Fig. 2). Recent work in the Tiki Basin[39], between the East Pacific Rise and the Marquesas, establishes the extent and nature of the western limb along 15° S. In both instances hydrothermal precipitates are being transported for thousands of kilometres before eventual deposition. Moreover the direction of transport is strongly dependent on latitude. The western lobe is exactly coincident with the ^{3}He plume recently identified by Lupton and Craig[40] (Fig. 3). This remarkable feature appears to originate as an intense ^{3}He maximum over the ridge axis at 15° S and propagates westwards at mid-depth with no discernable 'leakage' to the east.

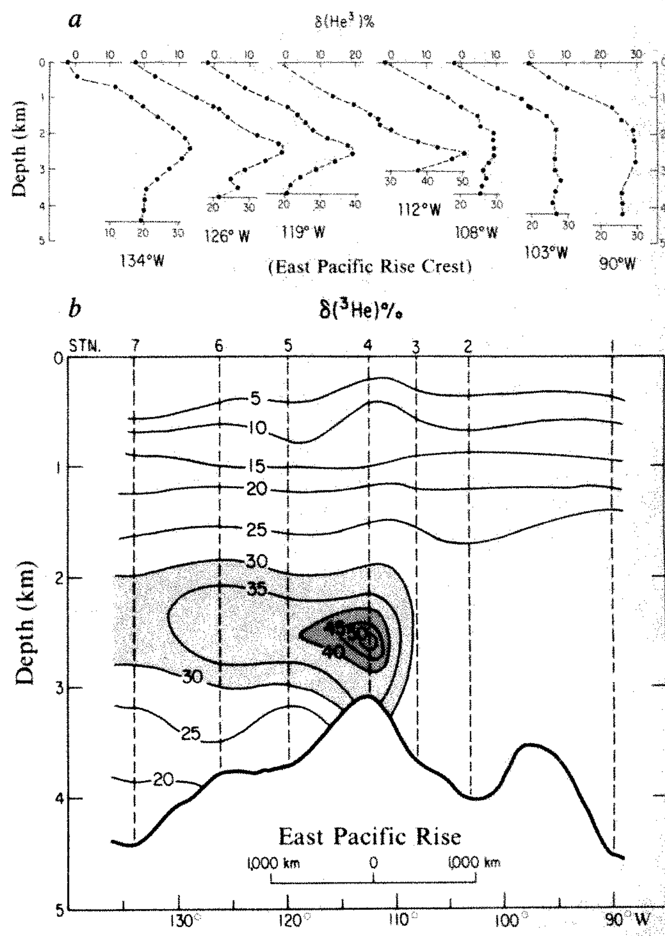


Fig. 3 δ^3He section along ~15° S in the south-east Pacific[40]. *a*, The individual profiles; *b*, the contoured values.

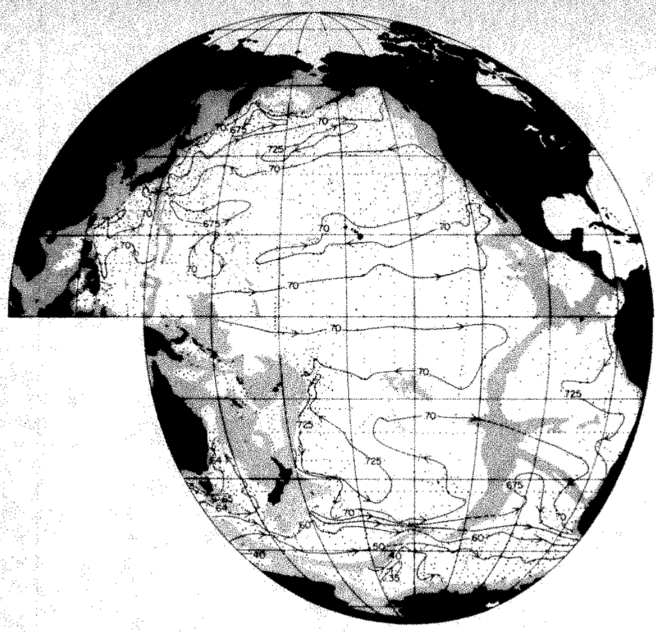


Fig. 4 Steric height at 2,000–3,500 db in dynamic metres ($10\ m^2\ s^{-2}$ or $10\ J\ kg^{-1}$) for the Pacific[42]. Shaded area is <3,500 m water depth.

These strongly developed asymmetries require a pronounced latitudinal zonality in the mid-depth circulation in the South Pacific, contrary to the predictions of the classical geostrophic models[41]. That this indeed exists has been dramatically demonstrated recently by Reid[42] who mapped the distributions of salinity, oxygen and the major nutrients on an isopycnal surface which is coincident with the ridge crest at mid-latitudes in the Pacific. He has also mapped the steric height at 2,000 db relative to 3,500 db (Fig. 4). Interpreting this measure of the relative geostrophic shear in terms of the mean flow directions (away from the western boundary) results in a circulation scheme consistent with the associated distributions of salinity and the non-conservative species[42]. Moreover comparison of the figures presented here shows the coincidence of the flow field with the distribution of metalliferous sediments and of ^3He to be exact.

The shear field in the South Pacific contains an anti-cyclonic gyre near 35–40° S. An eastwards flow leaves this gyre along 25° S in the east. There is a pronounced westwards flow along 15° S and an eastwards flow in the equatorial region. Flow over the basins east of the East Pacific Rise is generally southwards. This pattern is mirrored in the North Pacific (Fig. 4).

Of particular importance here is the circulation pattern in the region of the axis of the East Pacific Rise. Between about 5° and 15° S the flow is towards the south-southwest parallel to the rise crest. At 15–20° S the flow lines turn abruptly westwards and continue in that direction across the ocean. By 25° S the flow directions are reversed, passing east-southeast directly normal to the axis. In the extreme south-east Pacific, south of about 35° S the flow is to the north-east along the crest.

The exact concordance of the independently derived distributions of hydrothermal sediment and ^3He with the inferred flow is a remarkable demonstration of the importance of the mid-depth circulation in the dispersal of the hot spring effluents injected all along the East Pacific Rise. There are several consequences to this. First, and most obvious, the ^3He 'plume' along 15° S is but one section through a complex three-dimensional feature. The major axial anomaly results from the accumulation of ^3He injected along a flow-line almost 1,000 km in extent—the polewards flowing water above the ridge axis between 5° and 15° S. The plume is the expression of the dissipation of this anomaly along the westwards flow trajectory. The absence of any eastwards 'leakage' of ^3He is as expected. Since the extent of hydrothermal activity under the crestal station on the section (112° W) is not yet known it is not possible to separate the flow-integrating component in the ^3He maximum from any local injection. The very sharp spike of 50% δ^3He (Fig. 3) may reflect the immediate presence of a 21° N-type plume. Recent sampling on the East Pacific Rise has produced isolated ^3He anomalies of even greater magnitudes where the samplers were navigated to be immediately above known hydrothermal vents[43]. However, the broad maximum in δ^3He (35–45%) must be a regional rather than a local feature. Thus the spectacular section along 15° S does not require, of necessity, particularly vigorous hot spring activity on the axial zone beneath it. It does require that there be substantial injection of ^3He along the 5–15° S segment of the East Pacific Rise.

It should be possible to separate local from regional influences on the axial ^3He anomaly using parallel measurements of manganese[43]. As this element is rapidly removed by oxidation and sedimentation the regional $Mn/^3He$ anomaly ratio should be lower than any local one. The resolution of this approach requires much more data for its estimation as the rates of manganese removal are only crudely known[19].

Given the axial alignment of the flow between 5° and 15° S one might expect to see, in this segment, a progressive southwards increase in the abundance of metalliferous sediments accumulating as drift deposits. The data of Dymond[33] indicate this effect. While the spreading rate, a measure of the primary hydrothermal source function, increases from ~15 to 17.5 $cm\ yr^{-1}$ (ref. 44) between 5° and 15° S the percentage of hydrothermal sediments goes up by over 30% (Fig. 2) and the measured accumulation rates from 100 to >300 $mg\ cm^{-2}\ kyr^{-1}$ (ref. 31). While the latter data are sparse and the former biased by independent, systematic, latitudinal changes in the abundance of the other sedimentary components the trend is sufficiently suggestive and consistent to warrant more detailed study.

The systematics observed in the South Pacific have great potential importance in the validation of the estimates of the fluxes of hydrothermal effluents. The problem has two aspects; the generality of the computations for conservative species based on Galapagos and 21° N; the estimation of the general importance of sub-surface mixing—the average exit temperature—for the non-conservative elements. Components from either category which are unreactive in the water column at mid-depth will display an anomaly ratio to ^3He characteristic of the exit values at the hot springs. Sampling, if suitably located in the circulation field (Fig. 4), would provide values for these ratios integrated over particular segments of the ridge axis.

In the water column the concentration of a conservative species, C_A, is the sum of the background concentration, C_B, and an injected concentration, C_I:

$$C_A = C_B + C_I \tag{1}$$

The injected concentration is proportional to the injected ^3He:

$$C_I = R \times {}^3He_I \tag{2}$$

where $R = dC/d^3He$ for the discharging fluids. In terms of the 350 °C endmember:

$$R = \frac{C_{350} - C_A}{{}^3He_{350} - {}^3He_A} \tag{3}$$

Combining equations:

$$\frac{C_A - C_B}{C_A} = \left(\frac{C_{350}}{C_A} - 1\right)\frac{{}^3He_A - {}^3He_B}{{}^3He_{350} - {}^3He_A} \tag{4}$$

Recognizing that ${}^3He_{350} \gg {}^3He_A$ and that $\Delta^3He = ({}^3He_A/{}^3He_B) - 1$:

$$\frac{C_A - C_B}{C_A} = \frac{C_I}{C_A} = \left(\frac{C_{350}}{C_A} - 1\right)\frac{\Delta^3He}{{}^3He_{350}/{}^3He_B} \tag{5}$$

When C_{350} is known then the conservative value for C_I/C_A can be specified as a function of the ^3He anomaly: conversely when C_A and C_B are known at a given Δ^3He then C_{350} can be calculated.

The value for C_B is important both in determining the magnitude of the water column signature and in allowing accurate computation of the anomaly ratio. For ^{3}He, C_B is maintained at a low value because of its low solubility and relatively rapid escape from the atmosphere to space. For manganese efficient removal of the oxidized element from the water column produces the same effect. In the simplest case an average C_R can be established from work in the water column. Elements which are non-conservative in the sub-surface mixing regime will have exit concentrations which may be non-linear functions of temperature, depending on the details of the thermodynamic and kinetic responses of the precipitating phases to changing temperature, pH and so on[34]. This can be established empirically by sampling hot springs over the range of exit temperatures. It will be possible to do this at 21° N where individual hot spring occurrences have been observed spanning the range 20–350 °C. Non-conservative processes in the water column can be examined by sampling along flow lines, for example along 15° S. In principle therefore, the average fluxes from the ridge crest can be determined from work in the water column. However, a considerable amount of analytical development will be required for realizing this objective.

Conclusions

It seems that ridge crest hydrothermal activity is a pervasive and chemically relatively uniform phenomenon. The fluxes of the various elements injected into (or removed from) the ocean are substantial when compared with fluvial transport. The dispersal of this effluent is controlled by the global oceanic circulation at mid-depth. This effect allows integrated measures of effluent composition to be established from comparison of the water column anomaly ratios to ^{3}He at locations remote from the ridge axis. The sedimentary record of accumulation of precipitates of hydrothermal origin will therefore provide a history of variations in the circulation itself. This may be of some importance in palaeoclimatic studies. The internal consistency of the water column and sedimentary distributions of ridge crest effluents and the flow field validates the latter in a striking way. Maps of dynamic height for suitable depth ranges are therefore a valuable predictive tool in the planning of disposal of toxic or radioactive waste in the deep sea and in the exploration for and exploitation of mineral deposits of hydrothermal origin. For hot-spring organisms that have motile, benthic larvae it is likely that their dispersal is also controlled by the circulation field. The intriguing possibility therefore exists of a regional biogeography reflecting the latitudinal variations in the flow directions.

We thank Drs Lupton, Dymond and Reid for providing preprints and for stimulating discussions. J.M.E. thanks Dr C. Allègre for providing a place for work. This research is supported at MIT by grants from the NSF and International Decade of Ocean Exploration and from the US Office of Naval Research.

1. Lister, C. R. B. *Geophys. J. R. astr. Soc.* **26,** 515 (1972).
2. Elder, J. W. *Am. geophys. Un. Monogr.* No. 8, 211 (1965).
3. Bostrom, K. & Peterson, M. N. A. *Econ. Geol.* **61,** 1258 (1965).
4. Corliss, J. B. *J. geophys. Res.* **76,** 8128 (1971).
5. Thompson, G. *EOS* **54,** 1015 (1973).
6. Muehlenbacks, K. & Clayton, R. N. *Can. J. Earth Sci.* **9,** 172 (1972).
7. Gregory, R. T. & Taylor, H. P. *J. geophys. Res.* **86,** 2737 (1981).
8. Constantinou, G. & Govett, G. J. S. *Inst. Min. Metal. Trans.* **81,** B34 (1972).
9. Fleet, A. J. & Robertson, A. H. F. *J. geol. Soc. Lond.* **137,** 403 (1980).
10. Constantinou, G. & Govett, G. J. S. *Econ. Geol.* **68,** 843 (1973).
11. Dymond, J. *et al. Bull. geol. Soc. Am.* **84,** 3355 (1973).
12. Muehlenbacks, K. *Init. Rep. DSDP Leg* **37** (1976).
13. Clarke, W. B., Beg, M. A. & Craig, H. *Earth planet. Sci. Lett.* **6,** 213 (1969).
14. McKenzie, D. *J. geophys. Res.* **72,** 6261 (1967).
15. Wolery, T. J. & Sleep, N. H. *J. Geol.* **84,** 249 (1975).
16. Sclater, J. G., Parsons, B. & Jaupart, C. *J. geophys. Res.* **86,** 11, 535 (1981).
17. Bischoff, J. & Dickson, F. W. *Earth planet. Sci. Lett.* **25,** 385 (1975).
18. Lupton, J. E., Weiss, R. F. & Craig, H. *Nature* **267,** 603 (1977).
19. Weiss, R. F. *Earth planet. Sci. Lett.* **37,** 257 (1977).
20. Corliss, J. B. *et al. Science* **203,** 1073 (1979).
21. Crane, K. & Ballard, R. D. *J. geophys. Res.* **84,** 6011 (1979).
22. Jannasch, H. W. & Wirsen, C. O. *Bioscience* **29,** 592 (1979).
23. Edmond, *et al. Earth planet. Sci. Lett.* **46,** 1 (1979).
24. Edmond, J. M. *et al. Earth planet. Sci. Lett.* **46,** 19 (1979).
25. Edmond, J. M., Corliss, J. B. & Gordon, L. I. *Proc. 2nd Ewing Symp.* (ed. Talwani, M.) 383 (American Geophysics Union, 1979).
26. Fournier, R. O. & Rowe, J. J. *Am. J. Sci.* **264,** 685 (1966).
27. Jenkins, W. J., Edmond, J. M. & Corliss, J. B. *Nature* **272,** 156 (1978).
28. Craig, H., Clarke, W. B. & Beg, M. A. *Earth planet. Sci. Lett.* **26,** 125 (1975).
29. Cyamex Scientific Team. *Nature* **277,** 523 (1979).
30. Rise Project Group. *Science* **207,** 1421 (1980).
31. Edmond, J. M. *EOS* **61,** 992 (1980).
32. Craig, H., Welhan, J. A., Kim, K., Poreda, R. & Lupton, J. E. *EOS* **61,** 992 (1980).
33. Dymond, J. *Geol. Soc. Am. Mem.* (in the press).
34. McDuff, R. E. & Edmond, J. M. *Earth planet. Sci. Lett.* **57,** 117 (1982).
35. Albarede, F., Michard, A., Minster, J. F. & Michard, G. *TERRA Cog.* **1,** 50 (1981).
36. Brass, G. W. *Geochim. cosmochim. Acta* **40,** 721 (1976).
37. Spooner, E. T. C. *Earth planet. Sci. Lett.* **31,** 167 (1976).
38. Bostrom, K., Peterson, M. N. A., Joensuu, O. & Fisher, D. E. *J. geophys. Res.* **74,** 3261 (1969).
39. Hoffert, M., Karpoff, A. M., Schaaf, A. & Pautot, G. *La Genese des Nodules de Manganese* (ed. Lalou, C.) 101 (CNRS, Paris, 1979).
40. Lupton, J. E. & Craig, H. *Science* **214,** 13 (1981).
41. Stommel, H. & Arons, A. B. *Deep Sea Res.* **6,** 217 (1960).
42. Reid, J. L. *Evolution of Physical Oceanography* (eds Wunsch, C. & Warren, B.) 70 (MIT Press, Cambridge, 1981).
43. Merlivat, L., Dimon, B., Boulegue, J. & Francheteau, J. *TERRA Cog.* **1,** R11 (1981).
44. Anderson, R. N., McKenzie, D. & Sclater, J. G. *Earth planet. Sci. Lett.* **18,** 391 (1973).

ARTICLES

Motion of the Local Group of galaxies and isotropy of the Universe

L. Hart* & R. D. Davies

University of Manchester, Nuffield Radio Astronomy Laboratories, Jodrell Bank, Macclesfield, Cheshire SK11 9DL, UK

H I observations of 84 nearby Sbc spiral galaxies lying in the redshift range 1,000–5,500 km s^{-1} have been used to determine the motion of the Sun and the Local Group of galaxies relative to the backdrop of galaxies extending out to 70 Mpc. This study yields a Local Group velocity of 436±55 km s^{-1} towards $l = 264° \pm 18°$, $b = 45° \pm 12°$, in close agreement with the value inferred from the dipole anisotropy of the 3 K cosmic microwave background. The agreement shows that most of the dipole anisotropy is locally (extrinsically) induced and is not an intrinsic property of the Universe as has been implied. The component of the Local Group motion in the direction of the Virgo cluster is 410±55 km s^{-1}, a value similar to that derived recently by independent methods. This implies that the local value of the density of the Universe is 0.15–0.50 of that required to close the Universe.

* Present address: Instituto Argentino de Radioastronomia 1894 Villa Elisa (Prov. de Bs. As), Argentina.

0028-0836/82/200191—06$01.00

THE dipole anisotropy of amplitude ~3.5 mK in the 3 K cosmic microwave background has conventionally been interpreted as the result of the motion of the Sun at a velocity of ~400 km s^{-1} towards RA ~11.4 h, dec. ~5° (refs 1–3). When correction is made for the motion of the Sun relative to the centre of mass of the Local Group of galaxies, the resulting velocity of the Local Group becomes 550 km s^{-1} towards galactic coordinates $l=265°$, $b=35°$ (supergalactic coordinates $L=128°$, $B=-36°$). However, with the available observational data it is by no means certain that the dipole asymmetry in the cosmic microwave background is locally (that is, extrinsically) induced by the motion of the Sun and the Local Group. Dipole and higher order intrinsic asymmetries on the cosmic background have been proposed to explain the microwave observations. For example, Silk and Wilson[4] predict asymmetries arising from large-scale fluctuations in matter and radiation in the early Universe while Peebles[5] estimates that gravitational potential fluctuations induced by large-scale clustering could produce the observed large-scale asymmetries.

The suggestion that the microwave dipole asymmetry is largely intrinsic arises from the lack of agreement between the Local Group velocity inferred from the microwave measurements and that obtained from measurements of the motion of the Local Group relative to the backdrop of nearby galaxies. The most widely quoted and systematic search for this motion has been made by Rubin *et al.*[6,7] who found a motion for the Local Group of 450 km s^{-1} towards $l=163°$, $b=-11°$, a direction almost orthogonal to that given by the cosmic microwave asymmetry.

A genuine difference between the apparent motion of the Local Group relative to the microwave background and that relative to the backdrop of nearby galaxies has fundamental implications for cosmology. One possible consequence already alluded to is that the microwave background dipole asymmetry may be due to an intrinsic anisotropy[4,5,8]. Alternatively, if the microwave background is intrinsically isotropic and the dipole asymmetry is the result of the Local Group motion, then the implication is that the whole body of galaxies out to a radius of 90 Mpc is in systematic motion relative to the microwave background with a velocity of 750 km s^{-1}. Such large departures from a uniform Hubble flow are not, however, generally accepted[9]. The situation regarding a possible genuine difference between the Local Group motion determined by these two methods has not yet been resolved[10,11], particularly as different optical solutions for the Local Group motion have been published elsewhere[12–14].

We present here a new approach to deriving the Local Group motion relative to the backdrop of galaxies based on 21-cm wavelength H I observations alone, which obviates many of the difficulties in the optical determinations. Two independent H I data sets each give a Local Group motion in close agreement with that inferred from the dipole anisotropy in the 3K cosmic microwave background. The implications of these results for cosmology are discussed.

Rationale and observational material

The rationale behind our approach to measuring the Local Group motion was to use for distance measurement an H I standard candle independent of the optical methods which gave results disagreeing among themselves. The H I mass (proportional to the integrated H I flux density F_H) provides a standard candle and the half power velocity width $\Delta V_{1/2}$ of the integrated H I spectrum provides a third variable which indicates whether the galaxy is a giant or a dwarf. The H I integral is easily measured and calibrated in an all-sky survey and does not require the substantial corrections for galactic absorption, internal absorption, luminosity class, zenith angle, diameter and so on, which are applied to optical data. A detailed discussion of the observations and the selection criteria will be given elsewhere. However, we mention here that the dispersion in the H I mass of the sample of galaxies used in the present work is similar to the dispersion in the optical luminosity, when these properties are plotted against radius, luminosity class and so on. Ultimately we hope to use, H I, optical and IR data to obtain a composite distance indicator.

The objective of this programme of observations was to obtain a first-order solution for the local group motion using an optimal selection of galaxies whose motions were not seriously influenced by the gravitational potential of the Virgo cluster. Accordingly galaxies were chosen with velocities in the range 1,000–5,500 km s^{-1}; the small fraction of these galaxies likely to be most affected by the Virgo cluster were then eliminated from the list. A good sky coverage was achieved for galaxies in each velocity interval.

The galaxies used in this investigation were chosen from the list of 100 spiral galaxies of type Sbc (de Vaucouleurs type $T=4$) being studied in the λ21-cm line of neutral hydrogen at Jodrell Bank in a programme including optical, radio continuum and X-ray observations. These are essentially all the Sbc galaxies in the Second Reference Catalogue of galaxies[15] (RC2) for which luminosity classes have been assigned and for which the velocity, V_c, corrected for a solar motion of 300 km s^{-1} towards $l=90°$, $b=0°$ is $>1,000$ km s^{-1}. Most were of luminosity classes 1 to 4. H I observations have been made by Davies and Johnson (manuscript in preparation) of the integrated H I flux F_H, the full velocity width at half-power $\Delta V_{1/2}$ of the integrated profile and the systemic velocity V_c of each galaxy. To obtain a full coverage of the sky, the list was augmented by 12 galaxies at declinations south of dec. = −33° taken from a survey of bright southern galaxies by H. van Woerden *et al.* made at the Australian National Radio Astronomy Observatory (ANRAO), Parkes, New South Wales. The neutral hydrogen flux density scale of the latter survey was tied into the Jodrell Bank scale by comparing 18 galaxies

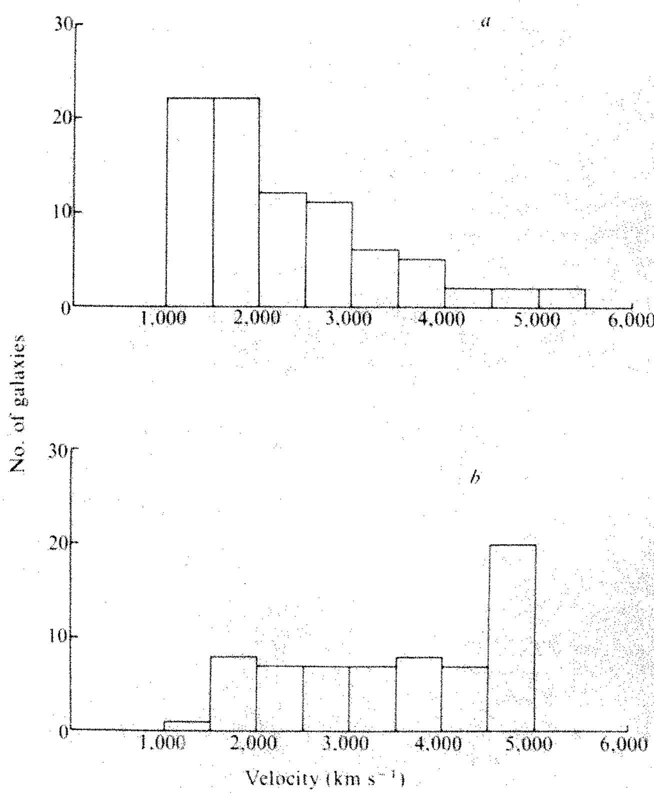


Fig. 1 The velocity distribution of galaxies used in the analysis of the Local Group motion. Velocities have been corrected by the standard solar motion towards the Local Group of 300 km s^{-1} towards $l=90°$, $b=0°$. Velocities $<1,000$ km s^{-1} have not been used. *a*, Sbc ($T=4$) galaxies from the Jodrell Bank sample supplemented by southern galaxies observed at ANRAO. *b*, Sc ($T=5$) galaxies from the Rubin *et al.* sample supplemented by southern galaxies; only galaxies with velocities $<5,000$ km s^{-1} are used.

Table 1 Observations of the motion of the Sun and the Local Group of galaxies

Method	Velocity of the Sun					Velocity of the Local Group*					
	$V_\odot$ (km s^{-1})	RA (h)	dec. (°)	l (°)	b (°)	V_{LG} (km s^{-1})	l (°)	b (°)	L (°)	B (°)	No. of galaxies
(1) H I fluxes: Sbc galaxies ($1{,}000 < V_c < 5{,}500$ km s^{-1}, $i > 30°$)	314 ±50	12.9 ±0.7	10 ±12	309 ±35	73 ±12	383 ±50	256 ±18	52 ±16	110 ±16	−27 ±14	84
(2) H I fluxes: Sbc galaxies (as for 1, but excluding high Virgo-centric velocities)	310 ±55	12.3 ±0.7	30 ±12	192 ±60	83 ±12	436 ±55	264 ±18	45 ±12	119 ±14	−29 ±12	78
(3) H I fluxes: Sc galaxies ($1{,}000 < V_c < 5{,}000$ km s^{-1})	419 ±62	10.2 ±0.6	19 ± 9	218 ±15	53 ± 9	580 ±62	245 ±18	35 ± 9	111 ±13	−45 ± 9	53
(4) Optical luminosity: Sbc galaxies	350 ±40	7.8 ±1.8	56 ±16	162 ±20	31 ±16	397 ±90	216 ±20	27 ±15	72 ±28	−55 ±16	79
(5) Microwave background† (ref. 2)	420 ±33	11.6 ±0.2	−12 ± 5	275 ± 7	46 ± 5	653 ±33	273 ± 6	27 ± 5	140 ± 6	−35 ± 5	
(6) Microwave background† (ref. 1)	401 ±60	11.2 ±0.5	19 ± 8	229 ±19	67 ±67	567 ±60	256 ± 9	41 ± 7	117 ± 8	−36 ± 7	
(7) Average of 2+3+5+6‡ (best estimates of motion)	368 ±50	11.5 ±0.5	15 ±12	243 ±30	68 ±12	546 ±70	261 ± 9	39 ± 7	122 ± 9	−35 ± 7	

* The solar motion relative to the Local Group of galaxies is taken to be 300 km s^{-1} towards $l = 90°$, $b = 0°$.
† The microwave background temperature is assumed to be 2.7 K: velocities are reduced by ~40 km s^{-1} for 3.0 K.
‡ Methods 2 and 3 are combined with weights 2 : 1.

common to the two lists. F_H and $\Delta V_{1/2}$ are accurate to better than 10% and V_c was accurate to 20 km s^{-1} or better for all the galaxies. The edge-on velocity width of each velocity profile, $\Delta V_{1/2}$ cosec $i = \Delta V_0$, was estimated using the inclination, i, of each galaxy derived from the axial ratio given in RC2 and an assumed intrinsic thickness-to-diameter ratio of 0.2.

Derivation of motion of the Local Group

This analysis is based on the use of the neutral hydrogen mass of a galaxy as a standard candle and using ΔV_0 as a third variable. The use of ΔV_0 in this analysis is similar to using the Fisher–Tully relation for optical distance determination.

In the analysis we take the observed velocity of a galaxy, V_c, corrected to the centre of the Local Group, to be composed of the following components

$$V_c = V_H + \Delta V_G + V_{pec}$$

where V_H is the velocity it would have in participating in the universal Hubble flow, ΔV_G is the component in the direction of the galaxy of the velocity of the Local Group relative to the backdrop of galaxies, and V_{pec} is the random velocity of the galaxy. V_H determines the true distance of the galaxy. We are interested to determine, from a series of expressions of this form, the vector V_G giving the velocity of the Local Group relative to the backdrop of Sbc galaxies. An estimate of the true distance of each galaxy, and hence V_H, is provided by its measured (apparent) H I flux density F_H where $F_H = M_H/D^2$, M_H is the H I mass and D is the distance.

Our first analysis follows the approach of Rubin *et al.*, where we express the difference in redshift distance and luminosity distance in terms of an H I Hubble modulus, HM, analogous to their optical HM in which

$$HM = 2 \log V_c - a \log \Delta V_0 + \log F_H \tag{1}$$

where a is the index of the calibration relation $M_H \propto \Delta V_0^a$. Then for each galaxy

$$\begin{aligned} -\mathbf{\Delta V}_G &= V_c - V_H - V_{pec} \\ &= V_c - 10^{1/2(\langle HM\rangle + a \log \Delta V_0 - \log F_H)} - V_{pec} \end{aligned} \tag{2}$$

We can use $\langle HM \rangle$, the average value of HM, and assume that V_{pec} is randomly distributed so that its expected value is zero to obtain a set of linear equations of the form

$$\begin{aligned} \mathbf{\Delta V}_{Gi} &= l_i X + m_i Y + n_i Z \\ &= V_c - 10^{1/2(\langle HM\rangle + a \log \Delta V_{0i} - \log F_{Hi})} \end{aligned} \tag{3}$$

which relates $\mathbf{\Delta V}_{Gi}$ to observed quantities where X, Y and Z are the components of the Local Group motion and l_i, m_i and n_i are the direction cosines of the ith galaxy. A least-squares solution of this over-determined set of equations gives an estimate of the vector $\mathbf{V}_G$.

A second and more rigorous statistical approach is to treat this as a well-defined non-linear minimization problem. We take the relation between M_H and ΔV_0 to be of the form

$$M_H = b\, \Delta V_0^a \tag{4}$$

M_H is related to the observed integrated flux density by $M_H = \text{const}\, F_H \cdot V_H^2$ where V_H is the Hubble velocity representing the true distance as defined above. Accordingly an estimate of V_H is

$$V_H = C\, \Delta V_0^{a/2} F_H^{1/2} \tag{5}$$

In the analysis V_{pec} is assumed to be normally distributed with a mean value of zero. The quantity to be minimized in determining the Local Group motion (X, Y, Z) is

$$\begin{aligned} \sum V_{pec\,i}^2 &= \sum [V_{ci} - (V_{Hi} + l_i X + m_i Y + n_i Z)]^2 \\ &= \sum [V_{ci} - c\, \Delta V_{0i}^{a/2} F_{Hi}^{-1/2} - (l_i X + m_i Y + n_i Z)]^2 \end{aligned} \tag{6}$$

The minimization is performed with respect to a, C, X, Y and Z. The solutions obtained by this more rigorous method agree well with those using the first method thereby indicating that the details of the analysis procedures are not too important for this set of data.

The main analysis described here is based on 84 Sbc galaxies with an inclination $i > 30°$ ($i = 0°$ for face-on galaxies) and $V_c > 1{,}000$ km s^{-1}. Figure 1*a* gives the velocity spread of the galaxies used in this analysis; the median velocity is 1,800 km s^{-1}. For galaxies more face-on than $i = 30°$ the inclination, and hence $\Delta V_{1/2}$ cosec i, becomes more uncertain. By making solutions for the Local Group motion with different cutoff values of i, we find that they are not strongly dependent on the cutoff value.

The motion found for the Local Group is 383 ± 50 km s^{-1} towards $l = 256° \pm 18°$, $b = +52° \pm 16°$ (supergalactic coordinates $L = 110° \pm 16°$, $B = -27° \pm 14°$). This solution corresponds

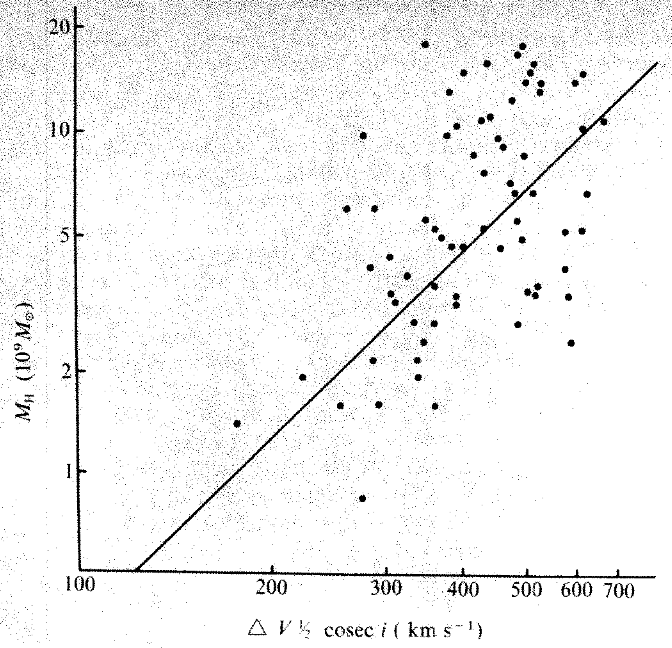


Fig. 2 The correlation between the neutral hydrogen mass M_H of Sbc ($T=4$) galaxies and their profile widths $\Delta V_{1/2}$ cosec i. $\Delta V_{1/2}$ is the width at half-power and i is the inclination of the galaxy. M_H is the neutral hydrogen luminosity indicator used in the solution for the Local Group motion. $M_H = F_H D^2$ where F_H is the integrated flux density and D is the distance in Mpc.

to a solar velocity of 314 ± 50 km s^{-1} towards RA = 12.9 h ± 0.7 h, dec. = 10° ± 12° (galactic coordinates l = 309° ± 35°, b = 73° ± 12°). The standard motion of the Sun relative to the Local Group of galaxies of 300 km s^{-1} towards l = 90°, b = 0° is used throughout.

Figure 2 shows the correlation between the neutral hydrogen luminosity of each galaxy, expressed as the total mass of neutral hydrogen, $M_H = F_H \cdot D^2$, plotted as a function of the observed profile width, ΔV_0, corrected for inclination. This correlation was obtained from the fitting procedures described above and yields a value of $a = 1.85 \pm 0.25$. Essentially each galaxy in Fig. 2 is placed at its 'true' redshift distance D based on the observed redshift and the derived value of the solar motion.

An indication of the validity of the analytical solution for the Local Group motion obtained here can be obtained by simply plotting as a function of supergalactic longitude L the difference between the observed redshift and the luminosity distance expressed as a redshift. Because the Local Group motion is in a direction close to the supergalactic plane this plot should show a sinusoidal variation of velocity difference with longitude. The amplitude plotted in Fig. 3 is 436 km s^{-1} directed towards L = 119°. The plot includes only the galaxies with $1{,}000 < V_c < 3{,}200$ km s^{-1} and supergalactic latitude $|B| < 50°$. The full solution in Table 1, of course, includes all the galaxies in the sample with $V_c > 1{,}000$ km s^{-1} and covers the whole sky.

It is relevant at this stage to comment on the possible effects on the solutions for the Local Group motion of Malmquist bias[6,16] in which the more distant galaxies in a magnitude limited sample tend to have higher intrinsic luminosity (or H I mass in our case). A plot of M_H against redshift shows only a weak Malmquist bias, largely because the galaxies in both the Sbc and Sc samples have a restricted luminosity class, corresponding to a restricted range of luminosity or H I mass. This latter property is illustrated by the small (~a factor of 2) range of ΔV_0 seen in Fig. 2. More important, the effects of any residual bias in the data are removed if there is a uniform distribution of galaxies in redshift around the sky; our galaxies satisfy this criterion. Further, the effects of Malmquist bias are also largely eliminated by the use of a third variable, ΔV_0, in our case. We are confident that the effects of Malmquist bias do not seriously affect our solution for the Local Group motion. This is confirmed by the fact that we obtain similar solutions for the Sbc and Sc galaxies and by using galaxies at different redshifts. These important matters will be discussed in more detail elsewhere.

The Virgo cluster of galaxies probably had a significant influence on the motion of the galaxies lying in the neighbourhood of the cluster; indeed we shall argue below that the Virgo cluster is responsible for the major part of the Local Group motion. Two solutions have been made using different criteria for excluding galaxies influenced by the Virgo cluster. In the first, members of the Virgo cluster are excluded; the Local Group is found to move at 434 km s^{-1} towards l = 262°, b = 52°. In the second, those galaxies with a line of sight component of Virgo-centric infall velocity greater than our own are eliminated; the Local Group is then found to move at 436 km s^{-1} towards l = 264°, b = 45°. We consider this to be our best estimate of the Local Group motion based on the H I observations of Sbc galaxies.

A value for the Local Group motion has also been derived from the optical data for the Sbc galaxies using an approach of minimum sophistication. The blue luminosities in RC2 were used as standard candles using the velocity width ΔV_0 as a third variable as for the H I analysis. This solution has lower precision than that obtained from the H I data; the solution is 397 ± 90 km s^{-1} towards l = 216° ± 20°, b = 27° ± 15°, a value similar to the H I result.

A completely independent data set is available for an analysis identical to that performed on our Sbc H I data. Rubin *et al.*[6] give H I parameters and plot H I spectra for a substantial fraction of their list of Sc galaxies. We have used all their galaxies with $1{,}000 < V_c < 5{,}000$ km s^{-1} and supplemented them with southern Sc galaxies from van Woerden *et al.* (in preparation) to provide a full sky coverage. Figure 1*b* shows the velocity distribution of this set of galaxies. The 53 Sc galaxies in this combined data set give a solution for the Local Group motion of 580 km s^{-1} towards l = 245°, b = 35°. This motion is consistent with the result obtained for the Sbc galaxies.

The results obtained from the present study are compared in Table 1 and Fig. 4 with the motion of the Sun and the Local Group relative to the microwave background derived on the assumption that the dipole anisotropy is entirely the result of solar motion. For a microwave background temperature of 2.7 K (ref. 17), the mean of the recent dipole determinations by Boughn *et al.*[2] and by Gorenstein and Smoot[1] implies a solar motion of 410 ± 45 km s^{-1} towards RA = 11.4 h ± 0.3 h and dec. = 3° ± 15° (l = 259°, b = 58°). The corresponding motion of the Local Group is 620 ± 50 km s^{-1} towards l = 265° ± 9°, b = 35° ± 7° (L = 128°, B = −36°). If the microwave background temperature is 3.0 K (ref. 18) instead of 2.7 K the solar velocity is reduced to 370 km s^{-1} towards l = 259°, b = 58° and the Local

Table 2 Component of the motion of the Local Group towards the Virgo cluster*

Method	Velocity (km s^{-1})	Ref.
(1) H I measurements of nearby Sbc and Sc galaxies with $1{,}000 < V_c < 5{,}000$ km s^{-1}	410±55	This paper
(2) 2.7 K cosmic microwave background, dipole anisotropy	450±50†	1, 2
(3) Four nearby clusters in IR with Fisher–Tully correction	480±75	23
(4) E galaxies in and around the Virgo cluster	470±75	24
(5) Mean value of methods 1, 2, 3, 4	447±40	

* The Virgo cluster is assumed to be centred at l = 280.4°, b = 74.6° (L = 102.2°, B = −3.0°). See ref. 26.

† Value for a background temperature of 2.7 K; the value is 430 km s^{-1} for a background temperature of 3.0 K.

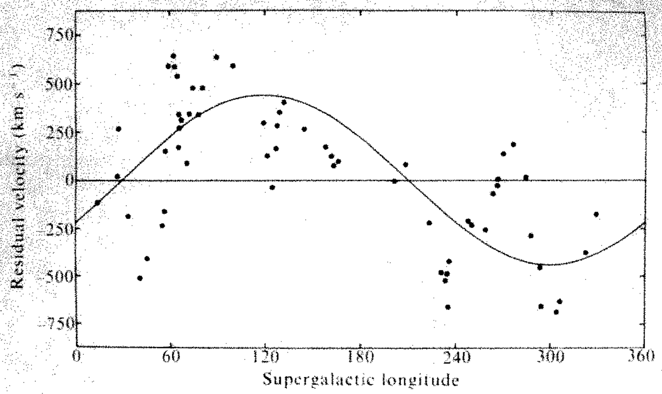


Fig. 3 The velocity residual between the luminosity distance of Sbc ($T=4$) galaxies estimated as a redshift and the observed redshift plotted as a function of supergalactic longitude, L. Only galaxies with $1{,}000 < V_c < 3{,}500$ km s^{-1} and within 50° of the supergalactic plane are included. The velocity residual shows a sinusoidal variation with L which is a consequence of the Local Group motion towards $L \sim 120°$. The sinewave plotted is that expected for a velocity of 436 km s^{-1} towards $L = 119°$, the solution found from a full analysis of all the Sbc galaxies; it is not a fit to plotted points alone—see text.

Group velocity reduces to 580 km s^{-1} towards $l = 266°$, $b = 32°$. The magnitude and direction of the solar motion determined from the H I observations of Sbc and Sc galaxies agree with the microwave dipole values to within two standard deviations (s.d.) of the combined errors.

The present results emphasize the disagreement between the Rubin *et al.* optical result[6,7] and that obtained from the microwave background measurements. The agreement found here between the determination based on H I data for the Rubin *et al.* galaxies and the other results in Table 1 confirms that the optical data base or the reduction procedures of Rubin *et al.* are in error[11,19]. Note that de Vaucouleurs and colleagues[12-14] find a solution for the solar motion calculated relative to the backdrop of spiral galaxies which is similar to our optical result. However, their solution is displaced by 4 s.d. from the statistically more accurate H I solutions which we believe are representative of the backdrop of galaxies since the H I sample is relatively free of contamination by galaxies whose velocities may be strongly affected by the Virgo cluster. The de Vaucouleurs galaxy samples on the other hand include ~30% with heliocentric velocities <1,000 km s^{-1} and a substantial number (~10%) of Virgo cluster members.

Consequences of Local Group motion

The close agreement between the velocity of the Local Group calculated relative to the backdrop of nearby galaxies and that derived from the 3 K cosmic microwave background dipole anisotropy argues strongly that the anisotropy is produced by the Local Group motion. This agreement is all the more significant as the velocity is directed to within 35° of the Virgo cluster ($l = 280°$, $b = 75°$, $L = 102°$, $B = -3°$) which is likely to have a major influence on the motion of the Local Group. The mean value of the motion derived from the two recent and independent microwave background results and from the two H I results is 546 ± 70 km s^{-1} towards $l = 261° \pm 9°$, $b = 39° \pm 7°$ ($L = 122° \pm 9°$, $B = -35° \pm 7°$). We consider that this is the best estimate of the Local Group motion available.

It is still possible that a fraction of the 3 K dipole anisotropy is intrinsic. An estimate of this intrinsic anisotropy can be derived from the velocity difference between the Local Group motion calculated relative to the backdrop of galaxies and that relative to the 3 K microwave background. The velocity difference is 130 ± 70 km s^{-1} for an assumed microwave background temperature of 2.7 K or 90 ± 65 km s^{-1} for 3.0 K. The corresponding difference amplitudes of the dipole component are 1.3 ± 0.7 and 0.9 ± 0.7 mK or 26–37% of the observed dipole anisotropy of ~3.5 mK.

Our present results therefore rule out the possibility that most of the cosmic background dipole anisotropy is intrinsic[4,5,20] and is due to an anisotropic cosmology or a large scale anisotropic matter distribution in the Universe. The relatively small velocity difference between the microwave background and the backdrop of galaxies implies much lower amplitude long-range galaxy correlations than suggested by Clutton-Brock and Peebles[8]. Further, our results raise some doubt about whether the tentative detections of some quadrupole components in the microwave background can be interpreted in terms of the anisotropic models of Silk and Wilson[4] or of Peebles[5]. Silk and Wilson predict quadrupole components which are 3×10^{-1} to 10^{-3} of the dipole component depending on the index of the power-law fluctuation spectrum; Peebles predicts a ratio of 3×10^{-1}. Our limit on the intrinsic dipole anisotropy apparently indicates that some reported quadrupole components are comparable with or greater than the intrinsic dipole component in disagreement with the above predictions.

Component of velocity towards the Virgo cluster

An evaluation of the component of the Local Group motion towards the Virgo cluster provides a means[21,22] of estimating the density of the Universe. Several estimates of this Virgocentric motion are already available; these will be compared and combined with the present determination of the Virgocentric motion.

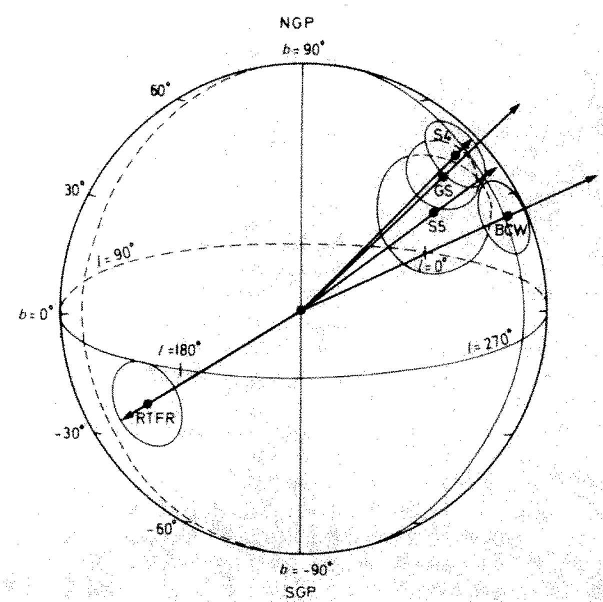


Fig. 4 Various solutions for the motion of the Local Group of galaxies plotted as vectors in galactic coordinates. Each solution is shown as a velocity with the appropriate length and with an error in angle described by a small circle on the surface of the celestial sphere. The vectors are coded as follows; S4, the Sbc ($T=4$) H I data of the present study; S5, the Sc ($T=5$) H I of the present study; GS, the microwave background dipole anisotropy from Gorenstein and Smoot; BCW, the microwave anisotropy from Boughn, Cheng and Wilkinson; RTFR, the Sbc galaxy optical data of Rubin *et al.* The agreement between the microwave background results and the H I results are evident. The Rubin *et al.* results are clearly discordant. The centre of the Virgo cluster lies at $l = 280°$, $b = 75°$ (from ref. 26).

The H I solution for the Sbc and Sc galaxies and for the microwave background give a Virgo-centric component of Local Group motion of ~ 450 km s^{-1} (Table 2). Another estimate of high precision (>5 s.d.) comes from work of Aaronson *et al.*[23] who used IR fluxes of individual galaxies corrected by the Fisher–Tully relation in four nearby clusters at redshifts in the range 4,000–6,000 km s^{-1}. Tonry and Davis[24] determined the Virgo-centric velocity from a study of E galaxies in and around the Virgo cluster. There is remarkably good agreement between these four high accuracy results.

Other estimates of the Virgo-centric motion[9,19,25] are available with lower quoted accuracy. Some of these refer to nearby galaxy samples which may themselves be influenced by the gravitational potential of the Virgo cluster. We consider that these results are not in serious agreement with the adopted Virgo-centric infall velocity of 447 ± 50 km s^{-1}.

Finally, we comment on the component of Local Group motion which is orthogonal to the direction of the Virgo cluster. The best estimate of the motion of the Local Group of 546 km s^{-1} towards $l = 261° \pm 9°$, $b = 39° \pm 7°$ (Table 1) is in a direction lying outside the errors of measurement of direction given by Sandage and Tammann[26] for the centre of luminosity of the Virgo cluster at $l = 280°$, $b = 75°$. Assuming that this is also the mass centre of the Virgo cluster, the difference in direction may be interpreted as the result of a component of motion in a direction orthogonal to the line of sight to the Virgo cluster. This motion is 320 ± 70 km s^{-1} towards $l \sim 254°$, $b \sim -14°$. Taken at its face value this motion is a 4.5 s.d. result for a Hubble flow departure in a velocity of 2,500 km s^{-1}, the mean of the Sbc and Sc galaxies. Such a motion provides a deviation $\Delta H_0/H_0$ in the Hubble flow of $320/2{,}500 = 0.13$ and is not at variance with the known departures[27] from the mean Hubble flow. However, the major component (300 km s^{-1}) of this orthogonal motion is directed towards the supergalactic plane and may be the result of collapse onto the flattened Virgo supercluster mass distribution[28]. The resultant motion in the Supergalactic plane perpendicular to the Virgo direction would be 100 km s^{-1} leading to $\Delta H_0/H_0 = 0.04$ for the random motion of the Local Group.

Virgo-centric motion and estimates of Ω and q_0

The evidence from the dipole anisotropy in the microwave background and the motion of the Sun relative to the backdrop of nearby galaxies strongly suggests that the Local Group is falling towards the centre of the Virgo cluster at a velocity of 447 ± 40 km s^{-1}. Such a departure from the Hubble flow of the Universe can be used to derive a local value for the deceleration parameter q_0 and for the ratio, Ω ($= 2q_0$ for Friedmann cosmologies), of the local density to the closure density of the Universe[19–22,24,29,30]. The calculation of q_0 and Ω requires an estimate of the density enhancement δ in the Virgo cluster relative to that in the smoothed-out Universe. Yahil *et al.*[29,30] find $\delta = 3.5 \pm 0.4$ while Davis *et al.*[31] find a value of 2.0 ± 0.3 from a different sample of galaxies. A detailed treatment[22,31] of the peculiar velocity v_p produced by a spherically symmetric density distribution gives, from a linear perturbation analysis,

$$\Omega = \left(\frac{3v_p}{\delta(v_p + v_{VC})}\right)^{1.67} \qquad (7)$$

where v_{VC} is the Virgo cluster velocity relative to the Local Group and is taken as 1,060 km s^{-1}. This yields a value of $\Omega = 0.10$–0.26 for $v_p = 447$ km s^{-1} and δ in the range 2–3.5. Using the results of a more exact non-linear calculation[31] we find $\Omega = 0.18 \pm 0.04$ to 0.39 ± 0.11 for this range of δ. Clearly the high value of the peculiar Virgo-centric velocity found in the present results and in the microwave background results implies a value of local density approaching the closure density, with $\Omega = 2q_0 = 0.15$–0.50 where the higher values correspond to the lower value of density enhancement ($\delta = 2$) in the Virgo cluster. Although this high value of Ω has been found for the local region of the Universe, essentially the local supercluster, there is no reason to believe that it is in any way different from the Universe at large. The infall velocity from which it is derived is the result of an acceleration acting over the age of the Universe.

Conclusion

The present observations demonstrate the use of a method of obtaining the statistical distances of spiral galaxies using H I data alone, namely H I fluxes and profile velocity widths. This method seems to be comparable in accuracy with related optical methods and has the advantage of requiring fewer correction factors. When applied to the determination of the motion of the Local Group of galaxies relative to the backdrop of nearby Sbc galaxies the method gives a velocity of 436 ± 55 km s^{-1} towards $l = 264° \pm 18°$, $b = 45° \pm 12°$, in agreement with the velocity implied by the dipole anisotropy in the 3 K cosmic microwave background.

The agreement between the motion of the Local Group determined relative to the nearby galaxies and that inferred from the microwave background is of fundamental significance. It implies that the major part at least of the dipole anisotropy is extrinsic, due to the Local Group motion, and is not intrinsic as implied in some cosmologies. The maximum intrinsic dipole anisotropy allowed by the combined data is ~ 1 mK, indicating that $\Delta T/T \leq 3 \times 10^{-4}$ for any intrinsic dipole anisotropy in the temperature distribution of the early ($z \sim 1{,}500$) Universe. The claimed quadrupole anisotropy[2] cannot then be the result of the processes envisaged by Silk and Wilson[4] and by Peebles[5].

The Virgo-centric velocity of the Local Group of galaxies is now well-determined from several independent methods. The inferred local values of q_0 and Ω depend on the assumed model of the mass density distribution in the Virgo cluster. In any case, high values of Ω (0.15–0.50) are implied by the observed Virgo-centric velocity. Such values, if applicable to the Universe as a whole, will produce a significant deceleration of the expansion of the Universe.

We thank Mr S. C. Johnson for discussions regarding the H I data base and Dr A. N. Lasenby regarding statistical techniques, also Professor H. van Woerden for supplying in advance of publication the southern H I data.

Received 21 September 1981; accepted 31 March 1982.

1. Gorenstein, M. V. & Smoot, G. F. *Astrophys. J.* **244,** 361–381 (1981).
2. Boughn, S. P., Cheng, E. S. & Wilkinson, D. T. *Astrophys. J. Lett.* **243,** L113–L117 (1981).
3. Fabbri, R., Guidi, I., Melchiorri, F. & Natale, V. *Phys. Rev. Lett.* **44,** 1563–1566 (1980).
4. Silk, J. & Wilson, M. L. *Astrophys. J. Lett.* **244,** L37–L41 (1981).
5. Peebles, P. J. E. *Astrophys. J. Lett.* **243,** L119–L122 (1981).
6. Rubin, V. C., Ford, W. K., Thonnard, N., Roberts, M. S. & Graham, J. A. *Astr. J.* **81,** 687–718 (1976).
7. Rubin, V. C., Thonnard, N., Ford, W. K. & Roberts, M. S. *Astr. J.* **81,** 719–737 (1976).
8. Clutton-Brock, M. & Peebles, P. J. E., *Astr. J.* **86,** 1115–1119 (1981).
9. Tammann, G. A., Sandage, A. & Yahil, A. *Phys. Scr.* **21,** 630–634 (1980).
10. Fall, S. M. & Jones, B. J. T. *Nature* **262,** 457–460 (1976).
11. Schechter, P. L. *Astr. J.* **82,** 569–576 (1977).
12. de Vaucouleurs, G. & Bollinger, G. *Astrophys. J.* **233,** 433–452 (1979).
13. de Vaucouleurs, G. & Peters, W. L. *Astrophys. J.* **248,** 395–407 (1981).
14. de Vaucouleurs, G., Peters, W. L., Bottinelli, L., Gouguenheim, L. & Paturel, G. *Astrophys. J.* **248,** 408–422 (1981).
15. de Vaucouleurs, G., de Vaucouleurs, A. & Corwin, H. *Second Reference Catalogue of Bright Galaxies* (University of Texas, Austin 1976).
16. Sandage, A., Tammann, G. A. & Yahil, A. *Astrophys. J.* **232,** 352–364 (1979).
17. Wilkinson, D. T. *Phys. Scr.* **21,** 606–609 (1980).
18. Richards, P. L. *Phys. Scr.* **21,** 610–613 (1980).
19. Yahil, A. *10th Texas Symp.*, (New York Academy of Science, in the press).
20. Wilson, M. L. & Silk, J. *Astrophys. J.* **243,** 14–25 (1981).
21. Silk, J. *Astrophys. J.* **193,** 525–527 (1974).
22. Peebles, P. J. E. *Astrophys. J.* **205,** 318–328 (1976).
23. Aaronson, M. *et al. Astrophys. J.* **239,** 12–37 (1980).
24. Tonry, J. L. & Davis, M. *Astrophys. J.* **246,** 680–695 (1981).
25. Schechter, P. L. *Astr. J.* **85,** 801–811 (1980).
26. Sandage, A. & Tammann, G. A. *Astrophys. J. Lett.* **207,** L1–L4 (1976).
27. Sandage, A. *Astrophys. J.* **202,** 563–582 (1975).
28. White, S. D. M. & Silk, J. *Astrophys. J.* **231,** 1–9 (1979).
29. Yahil, A., Sandage, A. & Tammann, G. A. *Phys. Scr.* **21,** 635–639 (1980).
30. Yahil, A., Sandage, A. & Tammann, G. A. *Astrophys. J.* **242,** 448–468 (1980).
31. Davis, M., Tonry, J., Huchra, J. & Latham, D. W. *Astrophys. J. Lett.* **238,** L113–L116 (1980).

Testing the theory of evolution by comparing phylogenetic trees constructed from five different protein sequences

David Penny*, L. R. Foulds† & M. D. Hendy‡

* Department of Botany and Zoology, and ‡ Department of Mathematics and Statistics, Massey University, Palmerston North, New Zealand
† Operations Research, University of Canterbury, Christchurch, New Zealand

The theory of evolution predicts that similar phylogenetic trees should be obtained from different sets of character data. We have tested this prediction using sequence data for 5 proteins from 11 species. Our results are consistent with the theory of evolution.

THE theory of evolution continues to be a focus for nearly all biological thought. Nevertheless, there have been doubts about the philosophical status of the theory, particularly on the extent to which it can be tested or falsified. The best known of these doubts have been expressed by the leading philosopher on scientific method, Karl Popper[1], who concluded that "darwinism is not a testable scientific theory, but a 'metaphysical research program'—a possible framework for testable scientific theories". Popper did not in any way reject evolution. He pointed out (ref. 1, p. 169) that "no serious competitor has come forward" and commented on "the strange similarity between my theory of the growth of knowledge and darwinism". In particular, he suggests that only one prediction is possible from darwinism: "Gradualism is thus, from a logical point of view, the central prediction of the theory. (It seems to me that it is the only prediction.)" (ref. 1, p. 172).

Popper has recently modified these criticisms[2], pointing out that criticisms by some authors were inconsistent, such as that natural selection is a tautology, and that it explains too much. A tautology explains nothing, so cannot simultaneously explain "too much". In addition, the suggestion was made that the existence of an evolutionary tree was falsifiable, but no reasoning was given for this new opinion.

These criticisms have aroused considerable interest[3-5]. Most of the discussion has, however, been qualitative, so it would be useful to be able to make quantitative tests. This is the purpose of the present article.

Popper makes the important distinction between the existence of an evolutionary tree (an evolutionary history) and the mechanism of evolution put forward to explain the processes that produced that history. Here we present a programme, applying graph theory[6,7], by which it is theoretically possible to refute the existence of an evolutionary tree.

Another prediction from evolution

It has been long been considered that protein sequence data contain evolutionary information[8]. In particular, the tree of minimal length (minimum number of mutations, maximum parsimony) makes no assumptions about the mechanism of evolution, and has been widely used as a model for evolutionary relationships[9-11]. Given any comparative data, irrespective of origin, one can construct trees of an evolutionary form, and hence a minimal tree must exist. So in general, finding a minimal tree for protein sequence data is not in itself independent evidence for the existence of an evolutionary tree.

However, another prediction can be made if there has been an evolutionary tree and if the maximum parsimony model is a good predictor of that tree. The prediction is that minimal trees with the same taxa should be similar, or 'congruent'[12], when constructed from different protein sequences. We need a measure of tree similarity having biological significance so that the values can be compared with the hypothesis that the trees are randomly selected. Such a measure is described below.

Our strategy is to take different protein sequences for a common set of taxa, find all the minimal (and near minimal) evolutionary trees and then compare them. Should the probability be high that these trees are unrelated, this would indicate that the protein sequences do not contain similar evolutionary information, and hence would contradict the existence of an evolutionary tree for those taxa.

We conclude that the existence of an evolutionary tree for these taxa is a falsifiable hypothesis, and therefore meets Popper's criteria for scientific theories. In addition, our methods allow us to identify a consensus tree which incorporates the most common features of all the near minimal trees.

Finding minimal evolutionary trees

It is easily shown that any minimal evolutionary tree can be represented by a binary tree. Using the double factorial notation (!!) there are $(2n-5)!! = 1\times3\times5\ldots\times(2n-5)$ unrooted binary trees spanning n sequences[13,14]. To determine the trees of maximum parsimony (minimal length), one must potentially consider a vast number of different tree topologies. But for all but a small number of sequences ($n \leq 8$), present day computers cannot consider them all[10].

Table 1 The frequency $p(m, n)$ of occurrence of $d(Ti, Tj) = m$ for binary trees spanning n taxa $4 \leq n \leq 11$

	$m=0$	2	4	6	8	10	12	14	16	$E(n)$
$n=4$	3.3×10^{-1}	6.7×10^{-1}								1.34
5	6.7×10^{-2}	2.7×10^{-1}	6.7×10^{-1}							3.20
6	9.5×10^{-3}	5.7×10^{-2}	2.4×10^{-1}	7.0×10^{-1}						5.24
7	1.1×10^{-3}	8.5×10^{-3}	4.6×10^{-2}	2.2×10^{-1}	7.3×10^{-1}					7.32
8	9.6×10^{-5}	9.6×10^{-4}	6.5×10^{-3}	3.8×10^{-2}	2.0×10^{-1}	7.5×10^{-1}				9.40
9	7.4×10^{-6}	8.9×10^{-5}	7.0×10^{-4}	4.8×10^{-3}	3.1×10^{-2}	1.9×10^{-1}	7.7×10^{-1}			11.46
10	4.9×10^{-7}	6.9×10^{-6}	6.2×10^{-5}	4.8×10^{-4}	3.6×10^{-3}	2.7×10^{-2}	1.8×10^{-1}	7.9×10^{-1}		13.52
11	2.9×10^{-8}	4.6×10^{-7}	4.7×10^{-6}	4.0×10^{-5}	3.3×10^{-4}	2.8×10^{-3}	2.3×10^{-2}	1.7×10^{-1}	8.1×10^{-1}	15.55

$E(n) = \Sigma\, mp(m, n)$ is the weighted mean or 'expected value' of $d(Ti, Tj)$ for randomly selected binary trees Ti, Tj spanning n taxa.

0028-0836/82/200197—06$01.00

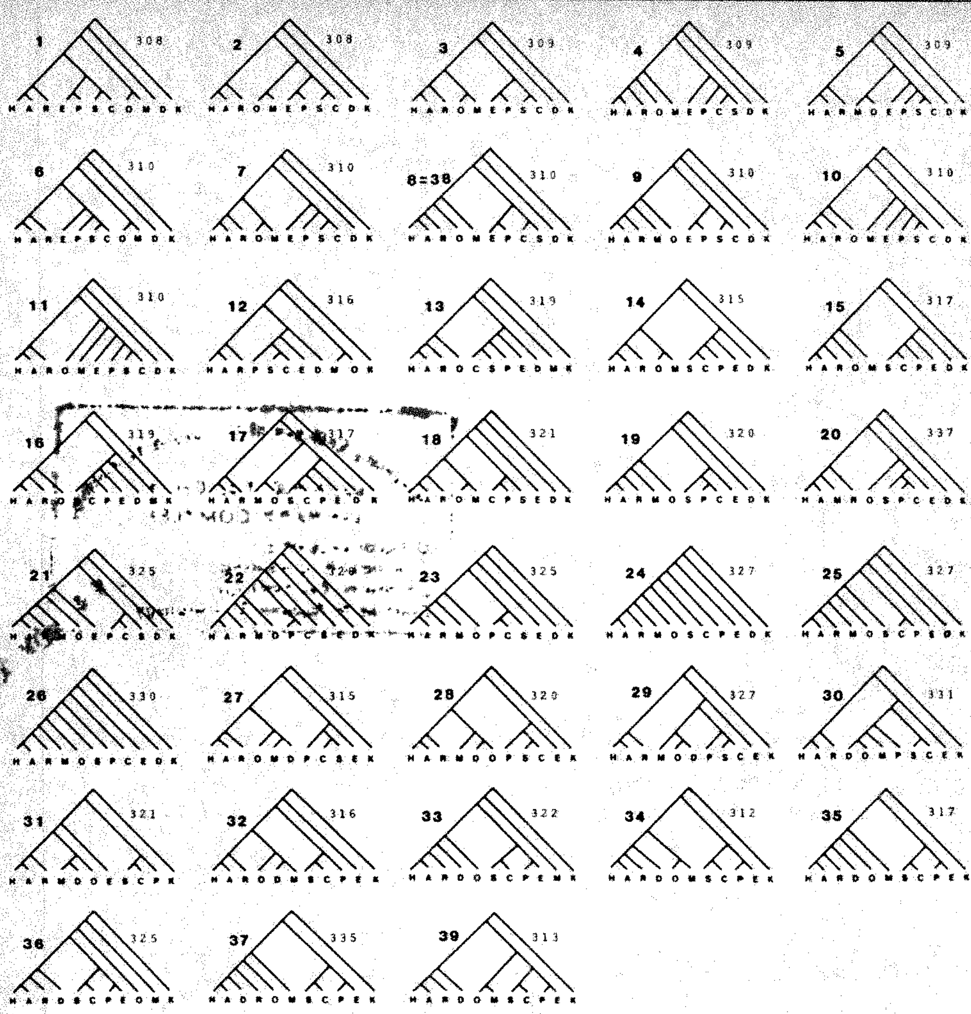


Fig. 1 The 39 minimal and near minimal trees spanning the 11 taxa given in Table 2. $T1$–$T11$ are generated from the complete sequences, $T12$–$T17$ from cytochrome *c*, $T18$ from fibrinopeptide A, $T19$–$T26$ from fibrinopeptide B, $T27$–$T32$ from haemoglobin *A*, and $T33$–$T39$ from haemoglobin *B*. The trees are arranged with their sequences in increasing lengths (see Table 3). The total number of mutations of each tree for the combined sequences is shown. The lengths of individual links are not shown, only the branching sequence is indicated. See Table 2 legend for definition of letter abbreviations.

We have shown that the problem of finding minimal length trees is an example of the Steiner problem in graphs[15,16] which makes no assumptions about the existence of an evolutionary history. Two recent techniques developed by the authors (refs 15–17 and D.P., in preparation) have enabled us to solve this for some larger sets of sequences (in one case a tree with 25 species has been proved minimal). These techniques rely on the inherent structural content of the data and are found not to work well on random data. This is an indication that there may be considerable tree-like structure within the sequence data. However, these observations are not easily quantifiable and the inability to use such methods on some data sets of this size ($8<n<25$) could not be taken as evidence for or against the existence of an evolutionary tree.

Here we have used a recently developed (ref. 17 and D.P., in preparation) 'branch and bound' method[7] which, for the data in Table 2, has found all minimal and near minimal trees.

Comparison of trees

Several methods have been proposed[18] that measure some features of the difference between a pair of trees, but it is important that any method chosen has biological relevance. Apart from satisfying a historical curiosity, a major application of an evolutionary tree is to assist in the classification of the taxa into groups at different levels. The set of taxa that are descendants of a given common ancestor will form a group of related taxa. In a strictly bifurcating (binary) tree spanning n taxa, there will be $n-2$ such non-trivial classes of the taxa. As in Robinson and Foulds[19], we measure the difference $d(T1, T2)$ of two trees $T1$ and $T2$, as the number of classes which are derived from $T1$ or $T2$, but not both. It is easily shown that this measure forms a metric on the space of all trees spanning these n taxa.

Given a non-directed tree T (one in which the root has not been specified), the removal of any internal link will partition the taxa into two subsets, one of which (depending on orientation) will be a group of related taxa. Waterman and Smith[20] have shown that the set of all such partitions uniquely defines T. Thus, $d(T1, T2)$ also represents the number of partitions of the taxa formed by deleting internal links, which differ in the two trees. This analysis is independent of whether or not the tree is a rooted tree.

Probabilities and tree comparisons

We can, for any specific tree T, determine the value of $d(T, T')$ for each of the $(2n-5)!!$ trees T'. Then the proportion of trees with $d(T, T')=m$ will represent the probability that a randomly selected tree T' is distance m from T.

For binary trees the number of internal links is $n-3$, so each tree defines $n-3$ partitions. If Ti, Tj have m partitions in common, $d(Ti, Tj)=2(n-3-m)$ which is even, and can range from 0 to $2n-6$. The value $d(Ti, Tj)=0$ can occur only for $i=j$, so the frequency of $d(Ti, Tj)=0$ is $1/N$, where $N=(2n-5)!!$. If $d(Ti, Tj)=2$ then there is only one link at which the two trees disagree. If we delete one internal link of Ti, there are two alternative ways of rejoining it to give values $d(Ti, Tj)=2$. This could occur at any of the $n-3$ internal links, so we find $d(Ti, Tj)=2$ with frequency $(2n-6)/N$.

The values of $d(Ti, Tj)$ over all pairs of binary trees spanning n species for $4\leq n\leq 11$ have been determined using recursive generating functions (D.P. and M.D.H., in preparation). These results are summarized in Table 1.

We can, for a given value of m, use these frequencies to estimate the probability of randomly selecting two trees Ti, Tj with $d(Ti, Tj)=m$. Referring to the case $n=11$, we would, for example, expect $d(Ti, Tj)=4$ to occur with probability

Table 2 Nucleotide sequences derived from the 5 polypeptides for 11 taxa

Taxon	Haemoglobin A	Haemoglobin B
R	CUCGGGGGGAUUACGGAAACUAGGAAAGGCAGCUGCAGAG	GCCACAAACACCCCCAUCUCCCGGCGGGACCAACCACAAGCUGACUAACAGCA
S	GAGGGGGGCAUUACGGGCAAAAGCCGGAGACUGUAACAGA	AACAGGCACGCUCCGAUCACGCGACGAUACAAAACGGCAGCGUGAUCAGUCAU
E	GACGGGGGGAUUGCGCAAACUAGCCGGGGACGUUAACGAG	GCGUGGCUUGCCCAGCUCAACAGGCCAUGAGCACCAGCGGCGUGAUACGACCA
K	GGGGAGGGGACAACACAAGCAAAACGGAAAAGCUGAAGGC	GCCAGAAACAGCAAACCAAAGAGCCGUGGCAAAAAAAAAAAAUAAUAAGAACG
M	GUCAGGAGGAUCACGGAAGAACCCCGGGGCAUCAGCCGAG	GCAUGGCUCGGCCCGAUCUCGCGGGACGACCAACCAUCGAAAUACCACGCGCU
O	CACAGAAGGAUUGAACAAAAAAGCCGGGACAAAUUCAGAA	GCGAAUCACGCCCAGAUCUCGCAACGCGAACAGACAAAGGCAUACUAACAGCA
D	CACAAAAGGCCAACGCAAAACACCCGGGGCAUGAACAGGC	GCCAGUCUCGGCCCACUCACCCAGGAAUACCAAAAAAAAGCUGACUAACAGCA
P	GGCGGGGCGAUUGAGCGAAAAAGCAGGGGCAUCUAACAAA	GCCUGGCUUGGCCCGAUCAAGCGGCGAUAACAAACAAAGGAGUGAUCCGUCCA
H	CACGGGGCGAUUACGGAAAAACCGAAAGGCAGCUGCAGAG	GCCACUCACGCCCCGAUCACCCGGCGGGACCGCCCAACGGCUGACUAACCGCA
C	GGGGGGGGCAUUACGGGCAAAAACCGGGGAAUCUAGCGAA	AACAGGCACGGUACGACAAAGCAACGAUACACAACGCAGGCGUGACCAGUCAU
A	CACGGGGCGCUUACGGAAAAACCGAAAGGCAGCUGCAGAG	GCCACUCACGCCCCGAUCACCCGGCGGGACCGCCCAACAGCUGACUAACCGCA

Taxon	Fibrinopeptides A and B		Cytochrome *c*
R	GAGGGAGGACCCG	AGAGAGAAGCCUAG	AAUUCAUCCCUAC
S	GCGUGGGGAACCG	AGGGCAACUCCUGA	GCAGCUUACACGA
E	CAGAGAGGAACAG	AGGACAAGUACUGA	GCAGCUAACACCA
K	CAAAGAGACAACG	AGGGUAAAGAGGGA	GCAGCUAAUCCGC
M	GAGGACAGAAAAC	AGCAAGUGAAUUAG	GCAGGUUACCCGC
O	GCGGGAAACAAAC	AGCGAGAGUUUCGA	GCAGGUUACCCAC
D	AUAAGAGGAAACG	AGGAUGUAGACGGA	GCAGCUUACACGC
P	GGCAAAGGAACCG	AGGCCAAGUCAGGA	GCAGCUUACACGA
H	GUGGGAGGACCCG	UAAGAGAGGUUUAG	AAUUCAUCUCUAC
C	GCCCAGGGACCCG	AGGCCAAGUGGUGG	GCAGCUUACACGA
A	GUGGGAGGACCCG	UAAGAGAGGUUUAG	AAUUCAUCUCUAC

R = Rhesus monkey (*Macaca mulatta*); S = sheep (*Ovis ammon*); E = horse (*Equus caballus*); K = kangaroo (*Macropus conguru* or *Macropus giganteus*); M = mouse (*Mus musculus* or *Rattus rattus*); O = rabbit (*Oryctolagus cuniculus*); D = dog (*Canis familiaris*); P = pig (*Sus scrufa*); H = human (*Homo sapiens*); C = cow (*Bos primogenius*); A = ape (*Pan troglodytes* or *Gorilla gorilla*).

4.7×10^{-6}. The weighted values, $E(n)$, given in Table 1 are the estimate of the expected value $(2n-6)$. Small values of $d(Ti, Tj)$ are very rare.

We have at this point established: (1) a method that can guarantee to find all minimal and near minimal trees, (2) a quantitative method of comparing trees and (3) a method of associating frequencies with the comparison of trees. The next stage is to apply these methods to comparative biological data.

Application to sequence data

When this study began there were polypeptide sequences available that were common to 11 taxa. The proteins are cytochrome *c*, haemoglobin A, haemoglobin B, fibrinopeptide A and the last 13 amino acids of fibrinopeptide B. The original data are from the *Atlas of Protein Sequence and Structure*[21], together with its supplements. There are cases, such as kangaroo, where two species of the same genus have been used (*Macropus conguru* and *Macropus giganteus*), but this does not affect the result. Table 2 lists the species. Our methods will work with more species and/or more sequences (for example, myoglobin, α-crystallin A, RNase), but we frequently find just one sequence is not available. There is an urgent need for more coordination in selecting sequences for analysis.

The protein sequences were converted to nucleotide sequences[22] (with haemoglobin B, some nucleotide sequences were available[23]). The sequence data were edited to remove invariant sites and other sites with no comparative information, as these have no effect on the structure of the phylogenetic trees (refs 10, 16 and D.P., in preparation). This left only a small number of sites, particularly for cytochrome *c*, as it shows little variation among these taxa. The final data are listed in Table 2, with 40 sites from haemoglobin A, 43 from haemoglobin B, 13 from fibrinopeptide A, 14 from fibrinopeptide B and 13 sites from cytochrome *c*.

Evolutionary tree construction

For 11 taxa there are 17!! = 34459425 unrooted binary trees to be considered. The branch and bound algorithm was applied to each of the five protein sequences individually, as well as to the combined sequences. Table 3 lists the numbers of trees close to minimal length for each of the five data sets and for the combined data. The 39 trees whose lengths are within 1.25% of the minimal lengths were selected for detailed comparisons and are illustrated in Fig. 1. In order that they should be presented as rooted evolutionary trees, the marsupial (kangaroo) was selected as determining the root or ancestor of the tree.

Tree comparisons

It can be seen from Fig. 1 that there are only two identical trees among these 39, $T8 = T38$. Using the comparison algorithm outlined above, we obtain the ${}_{39}C_2 = 741$ values of $d(Ti, Tj)$ ranging from 0 (1 value) to 14 (8 values) out of the maximum of 16, with a mean value of 7.57. Interpolating from the frequencies in Table 1, we would expect such a value to occur between a pair of random trees with probability 1.9×10^{-4}. There are, of course, 741 comparisons but they cannot be considered to be independent values. The average similarity for comparisons of trees from the same sequences is 4.1 and for comparisons of trees from different sequences is 8.33. The expected and observed values of m are as follows:

m =	0	2	4	6	8	10	12	14	16
Exp.	0	0	0	0	0	2	17	125	597
Obs.	1	53	87	163	200	145	84	8	0

Table 3 The minimal length and number of trees close to minimal length for each of the five sequences individually and combined

Sequence	Minimal length	min	(min + 1)	(min + 2)	(min + 3)	(min + 4)
Cytochrome *c*	17	6*	55	195	321	368
Fibrinopeptide A	29	1*	37	403	2,724	12,449
Fibrinopeptide B	36	8*	126	475	1,313	4,660
Haemoglobin A	89	1*	5*	26	143	400
Haemoglobin B	124	3*	4*	25	41	105
Combined sequences	308	2*	3*	6*	0*	8

* Those 39 trees with length ≤1.25% of the minimal. These 39 trees were used in the comparative analysis and are shown in Fig. 1.

Table 4 The average distance between trees derived from two sequences

	CS	Cc	FA	FB	HA	HB
CS	2.0/3.9	7.0	8.0	9.8	8.0	5.3
Cc	5.8	3.1	8.3	10.2	6.0	5.7
FA	6.9	8.3	—	4.8	10.0	11.3
FB	8.2	10.2	4.8	4.5	8.8	11.8
HA	8.4	7.6	10.3	11.2	—/4.7	6.7
HB	6.5	7.6	11.1	11.5	9.0	2.7/4.4

The values in the upper triangle are from comparisons among the 21 minimal trees, the values in the lower triangle are from comparisons among the 39 trees studied. CS = combined sequences, Cc = cytochrome *c*; FA, FB = fibrinopeptides A and B; HA, HB = haemoglobins A and B.

There is thus a strong divergence away from random towards the trees being very similar. Note that it is logically possible for the trees to have been more dissimilar than expected.

Table 4 gives the mean values of $d(Ti, Tj)$ between trees of different sequences, the upper values being obtained only from the minimal trees, and the lower values obtained from all 39 trees. The largest mean that occurs is 11.75 between the minimal length trees of fibrinopeptide B and haemoglobin B. This value would occur between a pair of random trees with probability 1.8×10^{-2}. All other values have probabilities less than this, ranging down to 1.1×10^{-5} for fibrinopeptides A and B.

Clearly, we can reject any idea that the trees from the different sequences are independent. The different protein sequences give trees that are markedly similar, showing a relationship between them that is consistent with the theory of evolution. This supports the theory but of course does not prove it; scientific theories are falsifiable but not provable[1,24].

Consensus tree

Of the 34459425 unrooted binary trees for 11 taxa, we have selected the 39 near minimal trees for further study. The minimal trees with the combined sequences are one estimate for the most likely tree from these data, but it is also possible to derive a 'consensus tree' which incorporates the most common features of the 39 trees.

We can, by using the partitions, find a particular tree T such that the sum $d(T, T1)+d(T, T2)+\cdots+d(T, T39)$ is minimal. In calculating the tree distances, we computed the number of times each particular partition occurred in a given tree. The nine most frequent partitions and the number of trees in which they occur are {H, A} (39); {H, A, R} (37); {S, C} (30); {S, C, P, E} (28); {S, C, P} (22); {K, D} (21); {H, A, R, O, M} (16); {E, P} (12); {O, M} (12) (see Table 2 legend for definition of abbreviations). For example, the first partition is {H, A} which is human and ape, and this occurs in all 39 trees. Each partition is defined by listing the smaller of the two subsets, and so in this case ({H, A}) the remaining nine taxa are in the other subset. The tenth most frequent partition occurs in only seven trees.

The partition {E, P} with frequency 12 is inconsistent with partition {S, C, P} with frequency 22. If we delete {E, P}, the remaining 8 partitions uniquely define the 8 links of a binary tree T spanning the 11 taxa. This tree is, in fact, the tree $T7$ in Fig. 1 of length 310 on the complete data and with an average distance of 5.5 to the other trees. We call this tree the consensus tree of the set of trees and it could be regarded as being representative of the set. It is the tree which will give the minimal sum of m values when comparisons are made with all 39 trees. It is a markedly better tree on this test than any other. Other consensus tree methods are available[25] but were less appropriate for this application.

The consensus tree does have the advantage over the two combined sequence minimal trees ($T1$, $T2$) in that in the ungulates, the horse (perissodactyl) is separated from the three artiodactyls (cow, sheep and pig). References to results of mammalian phylogeny from palaeontological evidence can be found elsewhere[26,27]. In McKenna's published scheme[26], lagomorphs (rabbit) would be the first branch after the marsupials. This does not occur in any of our 39 trees.

Conclusion

The general conclusions from the present work are that (1) it is possible to make falsifiable predictions from the hypothesis that species have been linked in the past by an evolutionary tree and (2) there is strong support from these five sequences for the theory of evolution. There may be exceptions where different sequences will lead to different trees as, for example, in the serial symbiosis theory[28]. Also, in pre-cellular evolution a network with circuits may be a better model than a tree[29]. Note also that this work has so far been confined to the question of the existence of an evolutionary tree and has not discussed the mechanism of evolution. The work can be extended by using criteria of optimality that assume particular mechanisms of evolution.

An interesting philosophical question would arise if the results of this work had falsified the prediction that the trees would be similar. Would this disprove the theory of evolution, or could it just mean that the sequences had changed so rapidly that they had lost all information about their early history, thus contradicting the hypothesis of Zuckerkandl and Pauling[8]? It could be argued that because proteins from different species can be aligned so readily, this in itself is independent evidence that the proteins retain evolutionary information. However, it is probably true that specific predictions from hypotheses, rather than the hypotheses themselves, are falsifiable. This idea is inherent in Popper's writing, but is more clearly expressed by Lakatos[30]. To this extent, we suggest that Popper's criticisms of evolutionary theory have shown incompleteness in the application of evolutionary theory, but at the same time evolutionary theory has helped clarify some inadequacies in Popper's model of the growth of knowledge.

Received 12 October 1981; accepted 16 March 1982.

1. Popper, K. *Unended Quest: An Intellectual Autobiography* (Fontana, London, 1976).
2. Popper, K. *Dialectica* **32,** 339–355 (1978).
3. Halstead, B. *New Scientist* **87,** 215–217 (1980).
4. Ruse, M. *New Scientist* **89,** 828–830 (1981).
5. Editorial *Nature* **290,** 75–76 (1981).
6. Harary, F. *Graph Theory* (Addison-Wesley, Reading, Massachusetts, 1969).
7. Carre, B. *Graphs and Networks* (Clarendon, Oxford, 1979).
8. Zuckerkandl, E. & Pauling, L. *J. theor. Biol.* **8,** 357–366 (1965).
9. Dayhoff, M. O. & Eck, R. V. *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Silver Springs, Maryland, 1966).
10. Fitch, W. M. *Am. Nat.* **111,** 223–257 (1977).
11. Goodman, M., Czelusniak, J., Moore, G. W. & Romero-Herrara, A. E. *Syst. Zool.* **28,** 132–163 (1979).
12. Mickevich, M. F. *Syst. Zool.* **27,** 143–158 (1978).
13. Cavalli-Sforza, L. L. & Edwards, A. W. F. *Evolution* **21,** 550–570 (1967).
14. Felsenstein, J. *Syst. Zool.* **27,** 27–33 (1978).
15. Hendy, M. D., Foulds, L. R. & Penny, D. *Math. Biosci.* **51,** 71–89 (1980).
16. Foulds, L. R. & Hendy, M. D. *J. molec. Evol.* **13,** 127–150 (1978).
17. Hendy, M. D. & Penny, D. *Math. Biosci.* **59** (in the press).
18. Smith, T. F. & Waterman, M. S. *Am. Math. Mon.* **87,** 552–553 (1980).
19. Robinson, D. F. & Foulds, L. R. *Springer Lect. Notes Math.* **748,** 119–126 (1979).
20. Waterman, M. S. & Smith, T. F. *J. theor. Biol.* **73,** 789–800 (1978).
21. Dayhoff, M. O. *Atlas of Protein Sequence and Structure 1972* (National Biomedical Research Foundation, Silver Springs, Maryland 1972).
22. Penny, D., Hendy, M. D. & Foulds, L. R. *Biochem. J.* **187,** 65–74 (1980).
23. van Ooyen, A. *et al. Science* **206,** 337–344 (1979).
24. Popper, K. R. *Objective Knowledge* (Oxford University Press, 1972).
25. Margush, T. & McMorris, F. R. *Bull. Math. Biol.* **43,** 239–244 (1981).
26. McKenna, M. C. in *Phylogeny of Primates* (eds Luckett, W. P. & Szalay, F. S.) 21–46 (Plenum, New York, 1975).
27. Szalay, F. S. in *Major Patterns in Vertebrate Evolution* (eds Hecht, M. K., Goody, P. C. & Hecht, B. M.) 315–374 (Plenum, New York, 1976).
28. Schwartz, R. M. & Dayhoff, M. O. *Science* **199,** 395–403 (1978).
29. Eigen, M & Winkler-Oswatitsch, R. *Naturwissenschaften* **68,** 217–228 (1981).
30. Lakatos, I. in *Method and Appraisal in Physical Science* (ed. Howsen, C.) 1–40 (Cambridge University Press, 1976).

DNA sequence of a foldback transposable element in *Drosophila*

S. Steven Potter

Department of Biology, Wesleyan University, Middletown, Connecticut 06457, USA

The bizarre structures of the Drosophila *foldback (FB) transposable elements are discussed here, and arguments favouring a transposase-mediated mobility are presented. The complete 4,089-base pair nucleotide sequence of transposable element FB4 reveals the full pattern of periodicity within the inverted repeats and the nature of the loop-inverted repeat junctions, as well as the coding potential of the loop sequence.*

TRANSPOSABLE elements are discrete units of DNA that are capable of occupying new genomic locations; they can covalently insert into different positions in the genome. Since their discovery[1], advances in technology have allowed the molecular isolation and characterization of a wide variety of transposable elements from bacteria and eukaryotes[2-9] (for reviews see refs 10, 11).

Here we present the complete nucleotide sequence (4,089 base pairs, bp) of the FB4 transposable element from *Drosophila melanogaster*. We have previously shown that the FB family of mobile sequences is distinct from the retrovirus-like transposable elements, such as *copia*[12,13]. The FB elements represent the only eukaryotic transposable elements discovered which have large inverted terminal repeats. Moreover, these FB terminal repeats vary in size from one element to the next, and many FB elements consist entirely of contiguous inverted terminal repeats, with no detectable separating DNA.

The FB4 nucleotide sequence is of interest for the following reasons. First, the data reveal the structural details of the periodicities found within the inverted repeats. Second, comparison of the two inverted repeats of FB4 shows differences that reflect the nature of some of the genetic events associated with these elements. Third, the loop sequence of FB4, separating the inverted repeats, is quite different from that found on any of the other cloned FB elements. Sequence analysis yields the structure of the loop–inverted repeat junctions and reveals the coding potential of the loop sequence.

The restriction map of FB4 and the sequencing strategy used are shown in Fig. 1. The complete base sequence of FB4 is presented in Fig. 2.

Inverted repeat structure

Each inverted repeat of FB4 is constructed primarily of small direct repeats. Near the distal termini of the inverted repeats there are multiple imperfect copies of the 10-bp sequence CGTTTGCCCA. These short repeats are, however, generally separated by large regions of more diverse DNA. At nucleotide 222 this 10-bp repeat is expanded to give 20 base pairs, CGTTTGCCCACCCTTTAAAA. This 20-bp repeat is also present in multiple imperfect copies, separated by variable regions of A+T-rich DNA. Starting at 499 bp from the end we find 31-bp contiguous tandem repeats which contain within them the 20-bp repeat. These 31-bp repeats continue for the ~500 bp remaining in the inverted repeats.

Figure 3 illustrates schematically this unusual sequence construction. The thick arrow represents one of the inverted repeats of FB4, and the shorter arrows the 10-bp repeats found near the distal termini. Figure 3 is only schematic, thus there are more copies of the various repeats than those shown.

Figure 4 illustrates the periodicity found within the tandem repeats of the 31-bp sequence. Not all copies of this repeat are identical. Five types of 31-bp repeat together make up a 155-bp (5×31) repeat that is itself tandemly repeated. In Fig. 4 the sequence from the proximal portions of the inverted repeats is shown with the 31-bp repeats in register and labelled according to type. The periodicity, with every fifth repeat being of the same type, is apparent.

This unusual sequence organization, with 10-bp repeats within 20-bp repeats within 31-bp repeats within 155-bp repeats within inverted repeats, leads us to speculate on the possible function. We propose the following scheme. Transposase levels in the eukaryotic cell, even in optimum conditions, may be extremely low, and the termini of an individual transposable element represent a minute fraction of the large genome. Therefore the rate-limiting step in the transposition process may be the recognition and binding of transposase to the ends of the transposable element, where the enzyme has its effect. The tandem repeats may serve to increase the size of the target

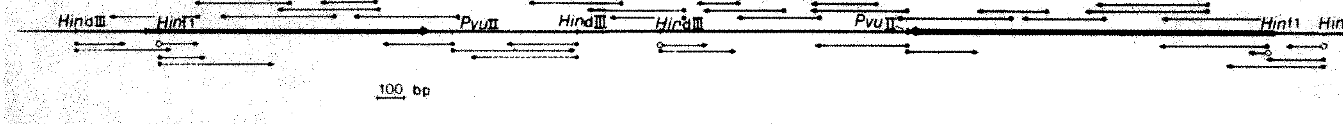


Fig. 1 Sequencing strategy. The arrows underneath represent sequencing experiments using the existing restriction recognition sites shown. Because of the repetitious nature of the inverted repeats there are relatively large regions of FB4 having few restriction sites. We therefore introduced new restriction sites by mutagenesis *in vitro*. The purified plasmid pDmFB4 (for a complete restriction map see Potter *et al.*[12]) was cleaved with *Xba*I and then treated for varying periods of time with the exonuclease *Bal*31. The reaction was monitored by gel electrophoresis. Synthetic polynucleotide linkers carrying *Xba*I recognition sites were blunt-end-ligated to the plasmid with T4 ligase. The DNA was then digested with *Xba*I, circularized by a second round of ligation, and used to transform *Escherichia coli* K-12 HB101. Ampicillin-resistant colonies were selected and screened by restriction analyses. The arrows above the map of FB4 represent sequencing experiments using the restriction sites introduced by *in vitro* mutagenesis. Enzymes and polynucleotide linkers were from New England Biolabs; isotope was from NEN. Sequencing was by the procedure of Maxam and Gilbert[28]. The five reactions—G, GA, CT, T and A—were used. End-labelling was performed as described previously[13,22,28].

0028-0836/82/200201—04$01.00

```
                                                                          HinfI
   1 AGCTCAAAGAAGCTGGGGTCGGAAAAAATCGAATTTTTGAAATTTGAAAGCTGCAATCGTTTGCCCATTTTTTGCCCATGTTTGCCCACCAATTAGTTTTT
 101 TTTGCCCACGTCCAGTTTTTGAGATATGGATTTTCGAAAAAGTTCGAAAATGTTCGAAAATCAAAAATTTCGCTTTTTTCAAATTTTTTTTTTTTTAAAT
 201 CGCAATAACATCGTTTGCCCACGTTTGCCCACCCTTTAGAATTTTGAAAAAATTTATACTTTAGAAAATATAAGGCTTTTAAGTTTACCTCGGTCTAATC
 301 AGAGAGTAAATCGTTTGCCCATCTCTTAAAACCAAATATTATCAACAAAAAACGTTTGCCCAACCATTATTATTAGTTTTTATCGTTTGCCCACCCTTTA
 401 AAAAACCTTTAACAAAATTTTTTTTTCGATTGCCCACACTTGAAATACAACCAATTTCGTTAGCCCACCTCTTCAAAATAAATATTTCCAATAAAAAACG
 501 TTTTCCCACCATTTAAAAATAAATAATTTCGATTGCCCATCCTTCAAAATTCATTTTAACGTTTGCCCACCCTTTAAAATTTGTTTTTTTCGTTTGCCCA
 601 CTCTTAAAACTAAATAATTTCGATTGCCCACCTTTTAAAACTAAATAATTTCGTTTGCCCATCCTTTAAAATTCATTTTTAACGTTTGCCCACCCTTTAA
 701 AAATAAATTATTTCGTTTGCCCACCCTTTAAAATTTGTTTTTTTCGTTTGCCCACTCTTAAAACTAAATAATTTCGATTGCCCACCTTTTAAAACTAAAT
 801 AATTTCGTTTGCCCATCCTTTAAAATTCATTTTAACGTTTGCCCACCCTTTAAAAATAAATTATTTCGTTTGCCCACCCTTTAAAAGTTTTTTTTTTTCG
 901 TTTGCCCACTCTTAAAACTAAATAATTTCGATTGCCCACCTTTTAAAACTAAATAATTTCGTTTGCCCATCCTTTAAAATTCATTTTTAACGTTTGCCCA
1001 CCCTTTAAAATTTGTTTTGTAAGATGTGGCGCCAATTCAGATATTTTAGGATCGGCGGATAGAAGCACTTACTTATATGATGATGATGAACATACATAGA
                 Pvu II
1101 CATAATATGTACAGCTGTGTTCAGAAAAATAGCAGTGCGAAGGAAACTAAGTAATACAAAGGTATTTTTCCATGTCCCTTTTCGGAATCGACTTTTTATT
1201 CCTCTTATTTTTGTTAAATGGAATGTGTAGATAGGGAAAAAAAGAAAATCCGGTCAGTTTTTCTTGTTATCCTTTTTTTATTTACATTCTTGAGCAAAAT
1301 CACAATTTTTAGGCTGTTCATAAGAATAGCAGTGTCTGGTTCTGACCAACGTAAAGTCCCGAAATGATCAATATTTTCTAAAAAGTGAGTTTGGTTAAGT
                                                                      CAT
1401 TAATTCGTATATTTAAAAGGCACAATAAATTAAAAAAATTAAAAAAATTTTATTTTAGTGGGTAGAGGACAGCACTACTCCCAGGGGAAAAGAATGTTAAT
         HindIII                  G-H                                                                Met
1501 TCTTAAGCTTAGAAAGGAAGGAAAAACATATAAGGACATTCAAAAAACCCTTAAATGTTCTGCCAAAATGGTATCCAATGCCATTAAATATAAATGGAAG
1601 CCCGAAAACCGTGGTACCAAACATAAAACCACAGATATAGAGGATCGACGCATTGTTTCTTACAGCAAAGTCTATCGTTTTGCATCCTTTAGGGACATAA
1701 AGTCTGAGCTGAACTTGGGAATCAGCGACGTTACTATTCGTAGACGACTACTGAATCAAAATTTCAGTGCGAGGAGTCCACGAAAGGTTCCCCTACCTAG
                                   HindIII
1801 CCCAAGGCATATTAAGGCAAGGTTAAGCTTAGCTAAAACCTACCTAAACTGGCCAGTCTCCAAATGGCGTAATATCCTTTGGACTGATGGGTCAAAAATC
1901 ATGCTATTTGGTGGAACTGGTTCACTACAGTATATCTGACGACCTCCAAACACGGAGTATCACCCAAAACACCCAGTGAAGACTTTCAATCACGGTGGAC
                                        End
2001 CTAAAATCATGGTATGGGCTTGTTTTTTTATAATGGTATGAGTCATGCTATGGATTATGATTTATGGTATTATAGACCAAAACGCATATGTAAATATAC
2101 TTAGTGATGTCTTATTGTCATATTCTGAATAAAATATACCCTTAAAATGGACATTCCAACAGGATAATGATCAGAAACGCAGATGTAAATCGGCTAAGAA
                                Poly A
2201 TAGGTTCACCCAAAATAGAATAGATGCAATGCCGTGGCAAGCACCACCTTCCCATTTAAACCCGATTGAAAACCTGTATGGGGACATTAAACAGTTTGTG
2301 TCGAAGAAGTCCCCGACGTCTAAGACTCAGATTTGGCAAGTTGTGCAGGATACATGGGCAAAAATTCCTCCCAAACCTTGCTAGGACTTGGTGGACTTCA
2401 TGCCGCGTGGGTGTAAGGCTGTGCTGGCTAACAAAGGCTATCCAGCCAAGTATTAGGCCCGAATTAACATATTAAAAAGAAAAACTAAGTTCGTTCTAGG
2501 TCAAGTTAAATTTTGTTACTATTTTTTCATAGCACTGCTATTTTATTGAACACCAGAATTTCTGCCTATTTATTGTTTTAATCTATATTTTCGAAACTAT
2601 TGAAGAAATAAAAGTGAAACATTTGTTAAATTGTTTGAAATCAAATACCTAATGATATTATTAAAAAAAAATTCCCATTAAAACTGTAAATCATAGGAAT
                                                    Pvu II
2701 TTTTTATCTTAAACTCTGAAGTCCAAAGCACTGCTATTATTCTGAACACAGCTGTACATAACGAAATAATTTATTTTTAAAGGGTGGCAAACCTTAAAA
2801 ATGAATTTTAAAGGATGGGCAATCGAAATTATTTAGTTTTAAGAGTGGGCAAACGAAAAAAACAAATTTTAAAGGGTGGGCAAACGAAATACTTTATTTT
2901 TAAAGGGTGGGCAAACGTTAAAAATGAATTTTAAAGGATGGGCAAACGAAATTATTTAGTTTTAAAAGGTGGGCAATCGAAATTATTTAGTTTTAAGAGT
3001 GGGCAAACGAAAAAAACAAATTTTAAAGGGTGGGCAAACGAAATAATTTATTTTTAAAGGGTGGGCAAACGTTAAAAATGAATTTTAAAGGATGGGCAAA
3101 CGAAATTATTTAGTTTTAAAAGGTGGGCAATCGAAATTATTTAGTTTTAAGAGTGGGCAAACGAAAAAAACAAATTTTAAAGGGTGGGCAAACGAAATAC
3201 TTTATTTTTAAAGGGTGGGCAAACGTTAAAAATGAATTTTAAAGGATGGGCAAACGAAATTATTTAGTTTTAAAAGGTGGGCAATCGAAATTATTTAGTT
3301 TTAAGAGTGGGCAAACGAAAAAAAAAAACTTTTAAAGGGTGGGCAAACGAAATAATTTATTTTTAAAGGGTGGGCAAACGTTAAAATGAATTTTAAAGGA
3401 TGGGCAAACGAAATTATTTAGTTTTAAAAGGTGGGCAATCGAAATTATTTAGTTTTAAGAGTAGGCAAACGAAAAAAACAAATTTTAAAGGGTGGGCAAA
3501 CGAAATAATTTATTTTTAAAGGGTGGGCAAACGTTAAAAATGAATTTTAAAGGATGGGCAAACGAAATTATTTAGTTTTAAATGGTGGGCAAACGTTTTT
3601 TTTTGGAAATTTTTATTTTAAAGAGGTGGGCTAACGAAATTGGGTGTATTTTAAATGTGGGCAATCGAAAAAAAATTTTTGTTAAAGGTTTTTTAAAGGG
3701 TGGGCAAACGATAAAAACTAATAATAATGGTTGGGCGTTTTTTGTTGATAATATTTGGTTTTAAGAGATGGGCAAACGATTTACTCTCTGATTAGACCGA
3801 GGTAAACTTAAAAGCCTTATATTTTCTAAAGTATAAATTTTTTCAAAATTCTAAAGGGTGGGCAAACGTGGGCAAACGATGTTATTGCGATTTAAAAAAA
3901 AAAAATTTTGAAAAAAGCGAAATTTTTGATTTTCGAAAATTTTCGAACTTTTTCGAAAATTCATATCTCAGAAACTGGACGTGGGCAAAAAAAACTAATT
                                       HinfI
4001 GGTGGGCAAACATGGGCAAAAAATGGGCATACGATTCCAGCTTTCAAATTTCAAAAATTCGATTTTTCCGACCCCAGCTTCTTTGAGCT 4089
```

Fig. 2 Nucleotide sequence of the FB4 transposable element. The complete base sequence of FB4 is shown, starting with the first nucleotide of the left inverted repeat. The various imperfect copies of the 10- and 20-bp repeats within the left inverted repeat are underlined, as are the punctuation signals surrounding the open reading frame of the loop and the 33-bp inverted terminal repeat of the loop. CAT, the CAT box; G–H, the Goldberg–Hogness box; poly(A), the polyadenylation signal.

recognized by the transposase. In a two-stage binding process, the transposase would first attach loosely but specifically at the tandem repeats and then move randomly (although a more directed movement is possible) until encountering the FB termini, where tighter binding and catalytic activity would occur.

This 'bind and slide' model, although conjectural, does provide testable predictions. For example, other factors being equal, an FB element having larger inverted repeats (and hence a larger transposase target size) will undergo more frequent transposition. In addition, the model suggests an affinity column transposase purification procedure, using the tandem repeats for selective binding.

There is a precedent for the proposed two-stage recognition. For example, the *lac* repressor apparently binds loosely to any DNA sequence, moves randomly along the DNA until it locates the specific recognition site, and then binds tightly[14].

Comparison of inverted repeats

Although the two inverted repeats of FB4 are very similar, there are, nevertheless, many single base differences and more dramatic differences involving entire copies of the 31-bp sequence. There are two 31-bp repeats missing in the left inverted repeat (Fig. 4) which are present at corresponding positions in the right inverted repeat. Differences of this sort could be generated by unequal cross-over events. Given the structure of the FB elements, both interchromosomal and intrachromosomal unequal cross-overs are possible. An intrachromosomal event, involving the two inverted repeats of a single FB4, would delete sequence from one inverted repeat and transfer it to the other. Because there are no extra copies of the 31-bp sequence at the corresponding positions on the right inverted repeat, we conclude that interchromosomal events generated these particular deletions.

Intrachromosomal recombinations between the inverted repeats of a single FB element would also result in an inversion of the loop sequence and part of the inverted repeats. Such inversions in analogous structures in *Salmonella*[15] and Mu phage[16] have been related to the regulation of gene expression. It is tempting to consider that such gene control mechanisms may also exist in eukaryotes.

Because the FB elements afford multiple possibilities for out-of-register pairing, they may represent 'hotspots' for normal homology-dependent recombination events. In this way they would contribute to genetic diversity by simply accelerating the regular recombination process at selected points along the chromosomes.

Further comparison of the two inverted repeats of FB4 reveals a missing 31-bp sequence in the right inverted repeat (Fig. 4). Note also that the right inverted repeat structure is significantly larger, with approximately two extra copies of the 155-bp repeat. Our restriction mapping data for several FB members suggest that such a difference in size of the 'inverted repeats' of a single FB element is common. Indeed the 'loop' sequences, as observed by electron microscopy, are apparently often the result of this size difference. That is, the 'extra' tandem repeats from one of the inverted repeats can form the loop.

The FB4 element, however, represents a striking exception in that the bulk of the relatively large loop appears by restriction mapping to be unrelated in sequence to the inverted repeats.

The FB4 loop

The 31-bp repeats terminate abruptly at the proximal junctions with the loop. This is in marked contrast to their gradual development from the distal ends, where 10-bp repeats become 20-bp repeats, which in turn become 31-bp repeats.

Directly abutting the proximal end of the right inverted terminal repeat is a 33-bp sequence (unrelated to the 31-bp repeat) that is also found in a nearly perfect (32/33 bp) inverted repeat near the other inverted repeat–loop junction. That is, a 33-bp sequence at the right inverted repeat–loop junction is also found, in an inverted orientation, 95 bp from the left inverted repeat–loop junction. The significance of this observation, if any, is unknown. The presence of such terminal repeats suggests that the loop itself may be a transposable element, and that FB4 is a composite structure, consisting of a mobile element (loop) within a mobile element (FB).

One of the signal characteristics of a transposable element is the presence of short direct flanking sequences that result from the duplication of target site DNA during the insertion process. The *Pvu*II recognition sites near the loop-inverted repeat boundaries are within 8-bp palindromes (ACAGCTGT) which can be regarded as either part of the full 33-bp inverted repeats of the loop, or as direct repeats flanking a shorter version of these inverted repeats (now only 20 bp long). Further support for the loop being a distinct transposable element is provided by preliminary results which indicate that these loop sequences are present at numerous genomic positions, where they are not always associated with FB elements (S.S.P., unpublished observations).

Alternatively, the loop may represent an integral and natural component of the FB4 element. This explanation avoids some of the difficult questions posed by the 'transposable element within a transposable element' model. For example, why is there a 95-bp space between the left inverted repeat–loop junction and the 33-bp 'terminal' repeat of the loop? And why did the loop transposable element insert into the centre of the FB element? (If it inserted within one of the inverted repeats, then a polarity switch would be seen in the tandem repeats of

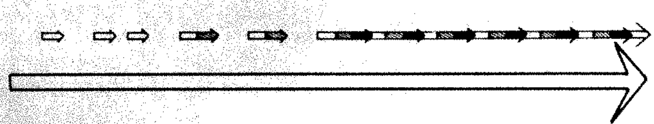


Fig. 3 Construction of the inverted repeats. The short arrows represent the 10-bp repeats, which expand to form 20-bp repeats which expand again to form 31-bp tandem repeats within the inverted repeat. There are many more copies of these various repeats than shown in this diagram.

Left	Right
1. CGTTTTCCCACCATTTAAAAATAAATAATTT	1. CGTTTGCCCACCATTTAAAACTAAATAATTT
2. CGATTGCCCATCCTTCAAAATTCATTTT_AA	2. CGTTTGCCCATCCTTTAAAATTCATTTTTAA
3. Absent	3. CGTTTGCCCACCCTTTAAAAATAAATTATTT
4. CGTTTGCCCACCCTTTAAAATTTGTTTTTTT	4. CGTTTGCCCACCCTTTAAAATTTGTTTTTTT
5. CGTTTGCCCACTCTTAAAACTAAATAATTT	5. CGTTTGCCTACTCTTAAAACTAAATAATTT
1. CGATTGCCCACCTTTTAAAACTAAATAATTT	1. CGATTGCCCACCTTTTAAAACTAAATAATTT
2. CGTTTGCCCATCCTTTAAAATTCATTTTTAA	2. CGTTTGCCCATCCTTTAAAATTCATTTT_AA
3. CGTTTGCCCACCCTTTAAAAATAAATTATTT	3. CGTTTGCCCACCCTTTAAAAATAAATTATTT
4. CGTTTGCCCACCCTTTAAAATTTGTTTTTTT	4. CGTTTGCCCACCCTTTAAAAGTTTTTTTTTT
5. CGTTTGCCCACTCTTAAAACTAAATAATTT	5. CGTTTGCCCACTCTTAAAACTAAATAATTT
1. CGATTGCCCACCTTTTAAAACTAAATAATTT	1. CGATTGCCCACCTTTTAAAACTAAATAATTT
2. CGTTTGCCCATCCTTTAAAATTCATTTT_AA	2. CGTTTGCCCATCCTTTAAAATTCATTTTTAA
3. CGTTTGCCCACCCTTTAAAAATAAATTATTT	3. CGTTTGCCCACCCTTTAAAAATAAAGTATTT
4. CGTTTGCCCACCCTTTAAAAGTTTTTTTTTT	4. CGTTTGCCCACCCTTTAAAATTTGTTTTTTT
5. CGTTTGCCCACTCTTAAAACTAAATAATTT	5. CGTTTGCCCACTCTTAAAACTAAATAATTT
1. CGATTGCCCACCTTTTAAAACTAAATAATTT	1. CGATTGCCCACCTTTTAAAACTAAATAATTT
2. CGTTTGCCCATCCTTTAAAATTCATTTTTAA	2. CGTTTGCCCATCCTTTAAAATTCATTTTTAA
3. Absent	3. CGTTTGCCCACCCTTTAAAAATAAATTATTT
4. CGTTTGCCCACCCTTTAAAATTTGTTTT	4. CGTTTGCCCACCCTTTAAAATTTGTTTTTTT
	5. CGTTTGCCCACTCTTAAAACTAAATAATTT
	1. CGATTGCCCACCTTTTAAAACTAAATAATTT
	2. CGTTTGCCCATCCTTTAAAATTCATTTTTAA
	3. CGTTTGCCCACCCTTTAAAAATAAAGTATTT
	4. CGTT[illegible]CCACCCTTTAAAATTTGTTTTTTT
	5. CGTTTGCCCACTCTTAAAACTAAATAATTT
	1. Absent
	2. CGATTGCCCATCCTTTAAAATTCATTTTTAA
	3. CGTTTGCCCACCCTTTAAAAATAAATTATTT
	CGT

Fig. 4 Periodicity of the 31-bp repeats. The sequences of the proximal portions of the inverted repeats are shown. Each type of 31-bp repeat is labelled by a number. The periodicity of types is apparent, although in three cases an expected copy of the 31-bp repeat is absent. To facilitate comparison, the reverse complement of the right inverted repeat is shown: the left inverted repeat sequence illustrated starts with nucleotide 499 and ends at position 1,018, while the right inverted repeat sequence shown starts with base 3,595 and ends at 2,761. Underlined bases deviate from the standard sequence of that repeat type. Underlined blank spaces represent a missing nucleotide that is normally present.

the other side of the FB element.) The integral component model is also attractive because it allows the FB4 element greater sequence complexity. The inverted repeats of the FB elements, although sometimes over 1 kb long, consist of repetitious sequences with frequent translational stop signals in all three possible reading frames, and it is therefore unlikely that they encode proteins. Where, then, is the transposase responsible for movement of these FB elements encoded? Loop sequences are obvious candidates, especially as bacterial transposons can carry transposase genes within their loops[17].

The loop sequence of FB4 contains a single large open reading frame, with the methionine codon beginning at nucleotide 1,494 and the next stop codon, TGA, beginning at nucleotide 1,938, allowing a maximum of 148 amino acids to be encoded. The consensus sequence GG$\frac{C}{T}$CAATCT, termed the CAT box, is often found upstream from eukaryotic genes. At the appropriate distance upstream from the coding region there is the sequence GATCAATAT which agrees with the CAT box for seven out of nine bases. There is also a good Goldberg–Hogness box and after the open reading frame, a perfect copy of the sequence AATAAA which is generally found 14–30 nucleotides upstream from the poly(A) tract of most mRNAs. It seems reasonable, therefore, to conclude that this open

reading frame sequence is a real gene, although the function of the gene product remains unknown.

It is also possible that other regions of FB4 encode small polypeptides, or that RNA splicing allows them to encode more substantial proteins.

Mechanism of transposition

Two distinct modes of duplication–transposition are now clear in outline, if not in detail. The bacterial transposons apparently move strictly via DNA intermediates, as proposed, for example, by the Shapiro model[18]. The finding of 'cointegrate' transposition intermediates[10], in which the two transposons are still linked, before the resolution event, strongly supports this general type of model.

The integrated proviral DNA copies of retrovirus RNA genomes undergo a more complicated duplication–transposition process. The proviral genome is transcribed into RNA which is packaged into a viral particle and eventually enters another cell, where it is reverse-transcribed[19], and a circular DNA intermediate having one or two copies of the long terminal repeats, is formed. This DNA circle can then integrate into the new host genome[20]. To generate retroviral intragenomic duplication–transposition events one need only eliminate the virion stage. The proviral transcript would be immediately reverse-transcribed, circularized and reintegrated at another chromosomal location.

The *copia*-like transposable elements in *Drosophila* are indistinguishable from integrated proviral retrovirus genomes. Structurally they both carry long direct terminal repeats[4,21,22] and the transcription patterns are very similar[23]. In addition, Flavell and Ish-Horowicz have shown[24] that *copia* elements give rise to circular DNA intermediates having one or two copies of the long direct terminal repeat, just as the retroviruses do. It therefore seems probable that the retroviruses and retrovirus-like transposable elements such as *copia* share a common duplication–transposition mechanism involving reverse transcription.

It has also been proposed that the human snRNA pseudogenes, with their dispersed and truncated organization, might result from a reverse transcription duplication–transposition process[25]. However, for these sequences, unlike the retroviruses which are designed to be reverse-transcribed, the process generally produces fragmentary versions of the original gene.

Jagadeeswaran *et al.*[26] have suggested that the human *alu* sequences could also duplicate and transpose by reverse transcription. Because these genes can be transcribed by RNA polymerase III, which recognizes internal promoters, the process is potentially 'explosive'. (The new *alu* genes can also be transcribed, giving rise to still more *alu* genes.) Furthermore, the *alu* transcripts would be capable of self priming for reverse transcriptase, giving rise to more faithful copies of the *alu* sequence.

Structurally the FB sequences are unique. They carry very large inverted terminal repeats, unlike anything found in a retrovirus, and do not seem to be designed for reverse transcription. Moreover, they have highly conserved termini, which argues against transposition by a less precise reverse transcription process, such as suggested for the *alu* and snRNA genes. Structurally the FB elements most closely resemble those bacterial transposons which have large inverted terminal repeats. As the bacterial elements move via transposase one can argue, by analogy, that the FB elements do also. Furthermore, as discussed previously, such a transposition mechanism can be invoked to help explain the unusual structure of the FB inverted repeats. The tandem repeats might serve to facilitate the binding of the low levels of transposase within the nucleoplasm. The retroviruses and retrovirus-like transposable elements, in contrast, rely on the much more abundant enzyme RNA polymerase to accomplish the onset of transposition. They therefore do not need to duplicate binding sequences.

Conclusion

The complete base sequence of FB4 reveals the full pattern of periodicity found within the inverted repeats. Near the distal termini there are occasional copies of a 10-bp sequence. Further along, this 10-bp sequence expands to 20 bp and this 20-bp repeat in turn expands to a 31-bp sequence which is tandemly repeated. There are five types of 31-bp repeat that together comprise a 155-bp repeat. This hierarchal repetition then terminates abruptly at the loop–inverted repeat junctions. Many differences exist between the left and right inverted repeats. The data suggest that these sequences are hotspots for unequal crossing-over.

The precise nature of the FB4 loop sequence remains uncertain. It is clear, however, that in this case most of the loop shows no significant homology with the inverted repeats. The presence of a gene within the loop is suggested by an open reading frame flanked by the correct punctuation signals.

It is intriguing that most FB elements cannot encode transposase. Such a dependence on transposase encoded elsewhere would seem hazardous to the transposable elements. Inactivation of the transposase gene(s) by deletion or mutation would render all the FB sequences immobile, at which point they might simply rest, accumulate mutations, and perhaps eventually be deleted themselves.

Finally, note that transposable elements interacting with transposase can generate a variety of chromosomal rearrangements, including inversions and deletions in contiguous sequences[18]. We have previously shown, by Southern blot experiments, the existence of several different DNA configurations within the vicinity of FB4 (ref. 12). The intriguing results of Goldberg *et al.*[27] also strongly suggest that two FB elements can cooperate to mobilize a large block of separating DNA. This further accentuates the many similarities found between the FB sequences and the bacterial transposable elements.

Received 5 February; accepted 26 March 1982.

1. McClintock, B. *Cold Spring Harb. Symp. quant. Biol.* **21**, 197–216 (1956).
2. Gafner, J. & Phillipsen, P. *Nature* **286**, 414–418 (1980).
3. Pardue, M. L. & Dawid, I. B. *Chromosoma* **83**, 29–43 (1981).
4. Potter, S. S., Brorein, W. J. Jr, Dunsmuir, P. & Rubin, G. M. *Cell* **17**, 415–427 (1979).
5. Roeder, G. S. & Fink, G. R. *Cell* **21**, 239–249 (1980).
6. Tchurikov, N. A., Zelentsova, E. S. & Georgiev, G. P. *Nucleic Acids Res.* **8**, 1243–1258 (1980).
7. Cameron, J. R., Loh, E. Y. & Davis, R. W. *Cell* **16**, 739–751 (1979).
8. Strathern, J. N., Spatola, E., McGill, C. & Hicks, J. B. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2839–2843 (1980).
9. Sakano, H., Huppi, K., Heinrich, G. & Tonegawa, S. *Nature* **280**, 288–294 (1979).
10. Calos, M. P. & Miller, J. M. *Cell* **20**, 579–595 (1980).
11. Spradling, A. C. & Rubin, G. M. R. *A. Rev. Genet.* **15** (in the press).
12. Potter, S., Truett, M., Phillips, M. & Maher, A. *Cell* **20**, 639–647 (1980).
13. Truett, M. A., Jones, R. A. & Potter, S. S. *Cell* **24**, 753–763 (1981).
14. Berg, O. G., Winter, R. & von Hippel, P. H. *Biochemistry* **20**, 6929–6948 (1981).
15. Zieg, J. & Simons, M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4196–4200 (1980).
16. Howe, M. H. *Nature* **271**, 608–610 (1978).
17. Heffron, F., McCarthy, B. J., Ohtsubo, H. & Ohtsubo, E. *Cell* **18**, 1153–1163 (1979).
18. Shapiro, J. A. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1933–1937 (1979).
19. Gilboa, E., Mitra, S. W., Goff, S. & Baltimore, D. *Cell* **18**, 1383–1396 (1979).
20. Shoemaker, C. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 3932–3936 (1980).
21. Finnegan, D. J., Rubin, G. M., Young, M. W. & Hogness, D. S. *Cold Spring Harb. Symp. quant. Biol.* **42**, 1053–1063 (1978).
22. Levis, R., Dunsmuir, P. & Rubin, G. M. *Cell* **21**, 581–588 (1980).
23. Flavell, A. J., Levis, R., Simon, M. A. & Rubin, G. M. *Nucleic Acids Res.* **9**, 6279–6291 (1982).
24. Flavell, A. J. & Ish-Horowicz, D. *Nature* **292**, 591–595 (1981).
25. Van Arsdell, S. W. *et al. Cell* **26**, 11–18 (1981).
26. Jagadeeswaran, P., Forget, B. G. & Weismann, S. M. *Cell* **26**, 141–142 (1981).
27. Goldberg, M. L., Paro, R. & Gehring, W. J. *EMBO J.* (in the press).
28. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499–560 (1980).

Avian sarcoma virus Y73 genome sequence and structural similarity of its transforming gene product to that of Rous sarcoma virus

N. Kitamura*, A. Kitamura, K. Toyoshima‡, Y. Hirayama & M. Yoshida†

Cancer Institute, Toshima-ku, Tokyo 170, Japan and ‡ Institute for Medical Sciences, University of Tokyo, Tokyo 108, Japan

From the complete nucleotide sequence of the genome of the avian sarcoma virus Y73, we have predicted amino acid sequence of p90$^{gag\text{-}yes}$, the product of the transforming gene. Contrary to previous evidence from molecular hybridization studies, p90$^{gag\text{-}yes}$ was found to have much homology with the transforming gene product p60^{src} of Rous sarcoma virus, suggesting that the cellular counterparts of the two (c-yes and c-src) originated from a common prototype sequence.

A GROUP of retroviruses, the acute leukaemia/sarcoma viruses, contains the transforming gene (v-*onc*), which is responsible for initiation and maintenance of cellular transformation[1]. More than 13 genes have been identified as v-*onc* and shown to be acquired from a normal cellular sequence as an integral part of their genome. Three distinct classes of transforming gene capable of inducing fibrosarcomas in chicken were defined among avian sarcoma viruses (ASV), that is, *src* in Rous sarcoma virus (RSV), *yes*[2,3] in Y73 and Esh sarcoma virus, and *fps*[4-6] in Fujinami and PRCII strains of ASV. These three transforming genes have no homology in nucleic acid hybridization[2-5], and are derived from the respective sequences in normal cells by recombination with viral sequences[2,7,8]. Furthermore, the cellular counterparts homologous with these transforming genes have been conserved during evolution[9] (M.Y., unpublished data) and expressed in normal cells[10], suggesting some that these genes perform some crucial role in normal cells.

In the light of these observations it is surprising that the transforming gene products of all these classes have protein kinase activity phosphorylating tyrosine on target proteins[11-14], and that this enzyme activity seems to be essential for cellular transformation[15,16]. Therefore, we have investigated the structure of these genes and the gene products in an attempt to understand not only the structural basis of the functions, but also the mechanism of biogenesis of the transforming genes.

The nucleotide sequence of *src* in Schmidt–Ruppin strain RSV (SR-RSV) has been established, and the function of the gene product (p60^{src}) and biogenesis of the *src* have been discussed on the basis of the nucleotide and amino acid sequences[17]. We have studied a recent isolate of ASV, Y73[18], which is defective in its replication and associated with leukosis virus (YAV) of subgroup A. The genome of Y73 is 26S RNA and contains *yes* as a transforming gene in its middle portion[19], which encodes polyprotein with a molecular weight of 90,000 (p90$^{gag\text{-}yes}$) consisting of a *gag*-gene derived part and Y73-specific portion. *In vitro* translation of 26S RNA[4] showed directly that p90$^{gag\text{-}yes}$ is coded by the Y73-specific RNA sequence. The association of protein kinase activity with p90$^{gag\text{-}yes}$ suggests that p90$^{gag\text{-}yes}$ is a transforming gene product.

To investigate the structure of p90$^{gag\text{-}yes}$, we determined the nucleotide sequence of the whole genome DNA. The sequence contains an uninterrupted reading frame that can code for p90$^{gag\text{-}yes}$ and the predicted amino acid sequence of p90$^{gag\text{-}yes}$ is strikingly similar to that of p60^{src}.

* Present address: Institute for Immunology, Kyoto University, Kyoto, Japan.
† To whom reprint requests should be addressed.

Sequencing strategy

For determination of the nucleotide sequence, we used the molecular cloning technique to amplify Y73 genome DNA. Closed circular DNA specific to Y73 was isolated from Y73-infected chick embryo cells by extraction of DNA by Hirt's procedure[20] followed by acid phenol extraction[21] and agarose gel electrophoresis. The purified DNA was linearized with restriction endonuclease *Sst*I and inserted into λ phage Charon 16A[22] at site of *Sst*I. Several clones containing the Y73-specific sequence were isolated and one of them, λY73-11A, was selected for sequencing, because it was shown by transfection assay to contain the complete sequence of Y73 RNA. Cotransfection of this cloned DNA together with cloned helper proviral DNA induced cellular transformation of chick cells and viral replication, the progeny virus was indistinguishable from the original Y73 virus. Details of the cloning the transfection will be reported elsewhere (manuscript in preparation).

Figure 1 shows a restriction map of clone λY73-11A insert containing two long terminal repeats (LTRs). The DNA fragments generated by digestion with *Hpa*II, *Hin*fI, and *Sau*3A were used for sequencing. The fragments were phosphorylated at the 5′ end with polynucleotide kinase and [γ-^{32}P]ATP, separated by either strand separation or further digestion with various restriction enzymes and sequenced by the method of Maxam and Gilbert[23]. Most of the sequences were confirmed by sequencing the complementary strand, or the fragments generated by digestions with different enzymes.

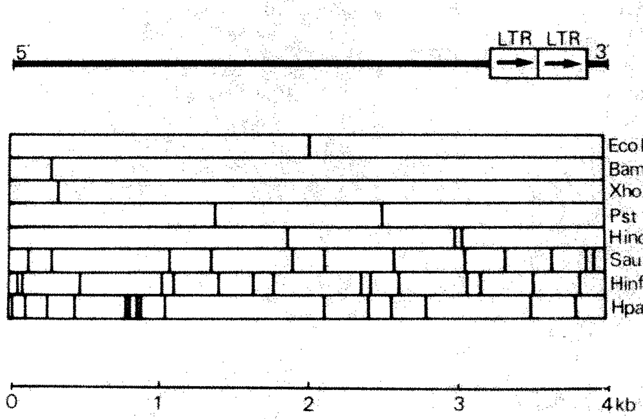


Fig. 1 Restriction map of Y73 genome cloned in λY73-11A. The circular DNA was linearized and inserted into λ Charon 16A vector at *Sst*I site and the cloned DNA contained two direct repeats of LTR.

0028-0836/82/200205—04$01.00

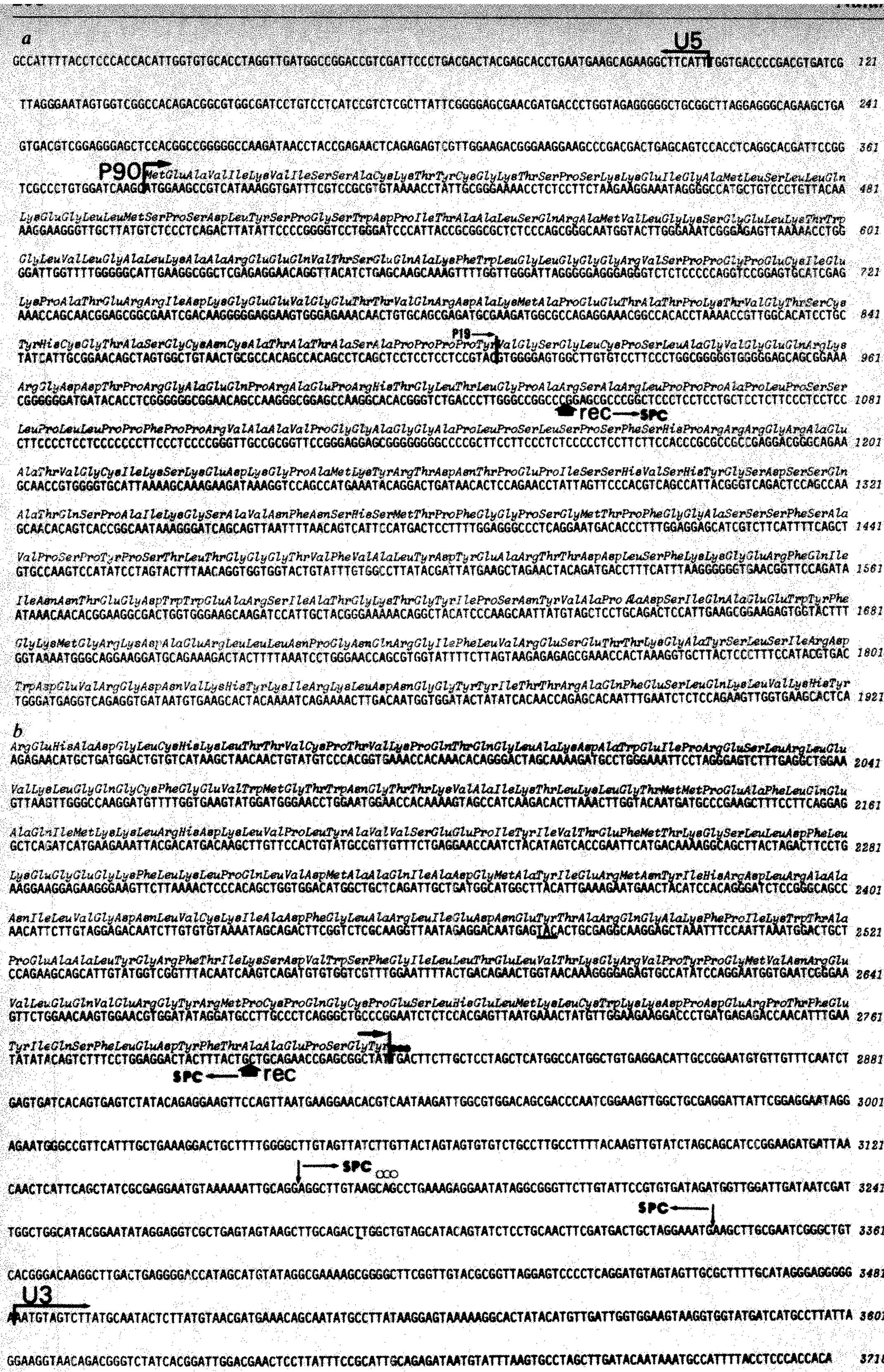


Fig. 2 Total nucleotide sequence of Y73 viral genome and the predicted amino acid sequence of p90$^{gag\text{-}yes}$. The complete sequence of the cloned DNA in λY73-11A was shown in the arrange of RNA genome from 5′ to 3′ direction. U5 and U3 indicate the unique sequence to 5′ and 3′ terminal regions. Thick solid arrows with 'rec' shows the predicted recombination sites between c-*yes* and YAV sequences, and 'spc' represents sequence which is not contained in RSV. The horizontal thick arrows indicate the initiation and termination sites for p90$^{gag\text{-}yes}$. The small arrow with p19 indicates a sequence homologous to the carboxy-terminus of p19 of Pr-RSV. The closed bar between nucleotides 2,477 and 2,479 shows the putative phosphoacceptor tyrosine contained in the same nonapeptide described by Neil *et al.*[26] and the small open circles between 3,174 and 3,176 show the position expected as termination codon for gp37, if present.

Nucleotide sequence

Previously, fingerprint analysis of Y73 26S RNA and YAV RNA[19] demonstrated that 26S RNA contained a specific sequence that represents the transforming gene of Y73. This specific sequence was mapped in the middle of the genome with sequences common to YAV RNA at both ends. The transforming gene product p90$^{gag\text{-}yes}$ encoded by the Y73-specific sequence was coprecipitated with antibody against p19, a *gag*-gene product, suggesting that p90$^{gag\text{-}yes}$ contained at least p19 in the amino-terminal portion[12].

The entire nucleotide sequence of the Y73 genome, shown in Fig. 2, consists of 3,718 nucleotides, confirming the previous data described above. The genome was significantly shorter than 4.8 kilobases, the value estimated previously by agarose gel electrophoresis of the genomic RNA.

The noncoding sequence at the 5′-terminus, including U5 and a primer tRNA binding site, was found to be homologous with that of the SR-RSV genome[31,32] (93% homology). The open reading frame, starting at position 380, contained the entire sequence coding for p19, which is located at the amino-

terminus of the *gag*-precursor, Pr76, indicating that AUG at position 380 is the initiation codon for p90$^{gag\text{-}yes}$. The proposed amino acid sequence of p19 was identical with that of p19 of Prague strain of RSV (D. Schwartz *et al.*, personal communication), except for the 2 amino acid insertions (positions 169 and 170) and 2 amino acid substitutions (positions 64 and 161). After the coding sequence for p19, 130 nucleotides were homologous with those of RSV and then there was a specific sequence. The sequence coding for p27, which follows after p19 in Pr76, was not present. Therefore, recombination between the viral sequence and the *yes*-related cellular sequence (c-*yes*) seems to have taken place in the sequence between those coding for p19 and p27. A pyrimidine-rich sequence was present after the putative recombination site, but, the significance of this sequence is not known. From the putative recombination site, the Y73-specific sequence extended to position 2,795. These specific 1,755 nucleotides are the only candidates for a coding sequence of the specific portion of p90$^{gag\text{-}yes}$. After this specific region, there was a sequence homologous to that coding for gp37 of SR-RSV[17], which is one of the *env*-gene products. Thus the recombination site at the 3′-side seemed to be within the sequence coding for gp37. Part of the sequence homologous to that coding for gp37 in SR-RSV was utilized for the carboxy-terminal peptide of p90$^{gag\text{-}yes}$ using a different frame from that for gp37. The UGA codon in the gp37 sequence was the predicted termination codon for p90$^{gag\text{-}yes}$.

Another 161 base sequence around the termination codon for gp37 is not homologous to that in PR-RSV (Schwartz *et al.*, personal communication), however, it seems too small to code any protein required for transformation. The origin of this specific sequence is not clear, but possibly the unique sequence was contained in, and originated from, the helper virus (YAV) genome. The common sequence at the 3′-end was homologous with that of SR-RSV[17,31] (87% homology) except for a 14 nucleotide deletion within the U3 region near the 5′ end. This deletion did not affect expression of the Y73 genome because the cloned DNA had transforming activity in transfection.

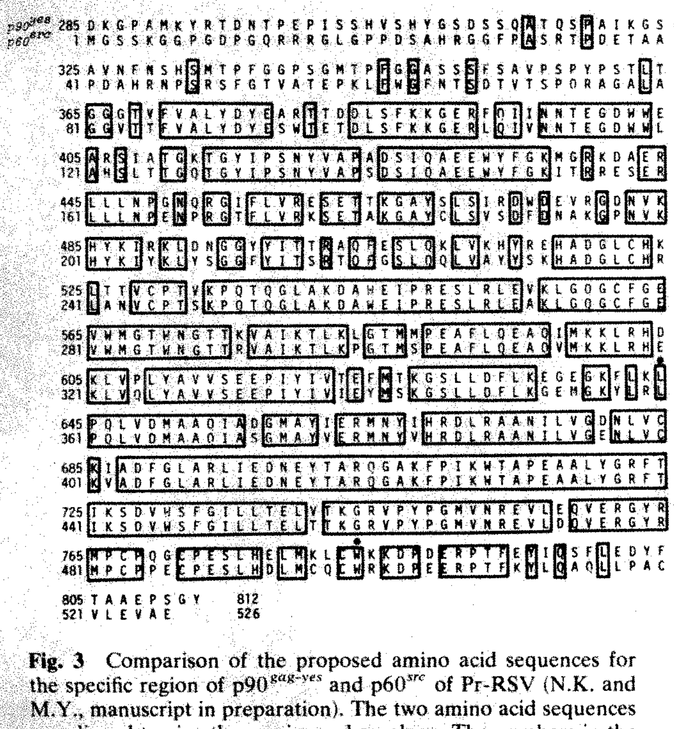


Fig. 3 Comparison of the proposed amino acid sequences for the specific region of p90$^{gag\text{-}yes}$ and p60^{src} of Pr-RSV (N.K. and M.Y., manuscript in preparation). The two amino acid sequences are aligned to give the maximum homology. The numbers in the left are the amino acid residues from the initiation methionine and area of identity are depicted by boxes. The nucleotide sequences between two marks (*) are compared in Fig. 4. A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.

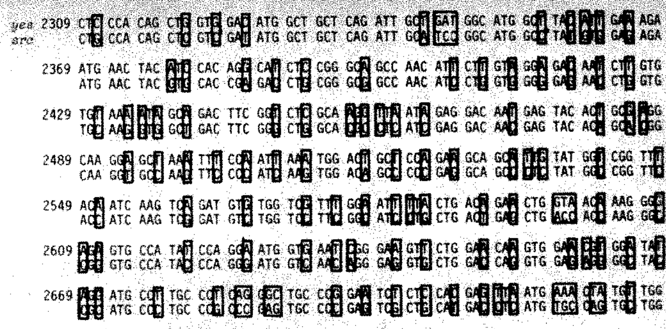


Fig. 4 Comparison between nucleotide sequences for the regions of *yes* and *src* of Pr-RSV coding for the highly homologous amino acid sequences. The nucleotide sequences corresponding to the amino acid sequences marked (*) in Fig. 3 are compared between *yes* and *src*, in which 93% homology was observed in amino acid sequences. Non-identical nucleotides are depicted by boxes.

Amino acid sequence of p90$^{gag\text{-}yes}$

Figure 2 also shows the predicted amino acid sequence of the gene product p90$^{gag\text{-}yes}$ of Y73. P90$^{gag\text{-}yes}$ consists of 812 amino acids and has a molecular weight of 88,697, which is in good agreement with the value of 90,000 estimated by SDS polyacrylamide gel electrophoresis[12]. As mentioned above, p19 was located at the amino acid terminus of the protein followed by a sequence unrelated to the RSV gene product.

Comparison of the amino acid sequence of part of p90$^{gag\text{-}yes}$ and p60^{src} is shown in Fig. 3. The amino acid sequence of p60^{src} was predicted by the nucleotide sequence of *src* in Pr-RSV, which was determined in our laboratory analysing molecularly cloned *Eco*RI fragment of the integrated proviral DNA (N.K. and M.Y., in preparation). Our sequence of *src* in Pr-RSV only differed in 12 nucleotides from that determined by Schwartz *et al.* (personal communication) but differed considerably from that of Czernilofsky *et al.* for SR-RSV, affecting a part of the reading frame for p60^{src}. Seventy-three out of 80 amino acids of the amino-terminus and 10 of the carboxy-terminus of p60^{src} were totally different from the corresponding regions of p90$^{gag\text{-}yes}$ when arranged to get maximum homology, but the other 436 amino acids of p60^{src} were largely homologous (82%) with those of p90$^{gag\text{-}yes}$. In particular 133 amino acids between positions 644 and 766 of p90$^{gag\text{-}yes}$ showed 93% homology with the corresponding region of p60^{src}.

Discussion

In earlier experiments[2,12], the Y73 specific cDNA sequence did not hybridize with *src* RNA. From this finding it was concluded that the transforming gene *yes* is distinct from *src* of RSV. However, our results show a large extent of homology in the amino acid sequences of the transforming gene products, p90$^{gag\text{-}yes}$ and p60^{src}. This discrepancy is due to the diversity of the genetic codes for each amino acid. Of 356 identical amino acids between positions 365 and 800, 221 utilized different codons from those for p60^{src} mainly owing to differences in the third base of the codons (Fig. 4). At the level of nucleotides, 31% of the 1,308 bases of the region are different in *src* and *yes*, and the longest homologous stretch is only 17 nucleotides long. These mismatches can explain the previous finding that cDNA of *yes* did not hybridize with *src* RNA under stringent conditions.

The highly homologous sequences of amino acids in p90$^{gag\text{-}yes}$ and p60^{src} are consistent with the fact that both gene products have protein kinase activity phosphorylating tyrosine[12,24,25]. Furthermore, these homologies strongly suggest that the two transforming genes were derived from a common prototype sequence. This was previously proposed by Neil *et al.*[26] from the finding of similar tryptic peptides containing phosphoaccep-

tor tyrosine in p60[src] and p90[gag-yes]. Our nucleotide sequence data clearly show homology over a wider range covering most of p60[src], thus supporting the proposal of Neil *et al.*[26]. Some similarity between p60[src] and the *mos* gene product of Mo-MSV was reported[27] and similarity was also found between two transforming gene products coded by Ha- and Ki-MSV[28].

Protein kinase activity phosphorylating tyrosine residues is a common feature of the transforming gene products of all avian sarcoma viruses so far tested. Therefore, it would be very interesting to know the amino acid sequences of the gene products of other avian sarcoma viruses such as Fujinami virus and PRCII.

As previously reported[2,8], there are DNA sequences homologous to those of *yes* and *src*, respectively in normal chick cells (c-*yes* and c-*src*), and these cellular sequences have been well conserved during evolution, being found in fish and in humans, suggesting some important function for these genes[9]. Therefore, the similarity of p90[gag-yes] and p60[src] at a protein level implies similarity of the gene products coded by c-*yes* and c-*src*, and suggests that c-*yes* and c-*src* diverged from a common prototype sequence at a very early stage of vertebrate evolution.

C-*src* was shown to be expressed in protein p60[c-src] in normal cells[10,11] and c-*yes* was shown to be transcribed into mRNA (M.Y., unpublished data). We have no data that could explain why different genes coding for proteins with similar structure and function are conserved and expressed in normal cells, although some speculations are possible. For example, expression of these cellular genes might depend on different types of differentiation or differences in the cellular distributions, and thus functions, of the proteins.

The structural similarity between p90[gag-yes] and p60[src] does not necessarily imply that the mechanisms of cellular transformation by these viruses are the same. Similarity in the carboxy-terminal domain could explain the similar protein kinase activity[29], but, differences in the amino-terminal domain suggest a different distribution of the proteins within cells, and thus that they phosphorylate different target proteins, since the amino-terminal domain of p60[src] is reported to be involved in anchorage of the protein to the plasma membrane[30].

We thank Dr Haruo Sugano for his useful discussions. This work was partly supported by Grant-in-Aid for Cancer Research from the Ministry of Education, Science and Culture of Japan.

Received 11 January; accepted 24 March, 1982.

1. Bishop, J. M. *A. Rev. Biochem.* **47**, 35–88 (1978).
2. Yoshida, M., Kawai, S. & Toyoshima, K. *Nature* **287**, 653–654 (1980).
3. Neil, J. C., Breitman, M. L. & Vogt, P. K. *Virology* **108**, 98–110 (1981).
4. Hanafusa, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 3009–3013 (1980).
5. Lee, W.-H. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 2018–2022 (1980).
6. Gysdael, J., Neil, J. C. & Vogt, P. K. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2611–2615 (1981).
7. Shibuya, M., Hanafusa, T., Hanafusa, H. & Stephenson, J. R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6536–6540 (1980).
8. Stehelin, D., Varmus, H. E., Bishop, J. M. & Vogt, P. K. *Nature* **260**, 170–173 (1976).
9. Spector, D., Varmus, H. E. & Bishop, J. M. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4102–4106 (1978).
10. Spector, D., Baker, B., Varmus, H. E. & Bishop, J. M. *Cell* **13**, 381–386 (1978).
11. Collet, M. S. & Erikson, R. L. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2021–2024 (1978).
12. Kawai, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 6199–6203 (1980).
13. Feldman, R. A., Hanafusa, G. & Hanafusa, H. *Cell* **22**, 757–765 (1980).
14. Neil, J. C., Ghysdael, J. & Vogt, P. K. *Virology* **109**, 223–228 (1981).
15. Levinson, A. D., Opperman, H., Levinthow, L., Varmus, H. E. & Bishop, J. M. *Cell* **15**, 561–572 (1978).
16. Powson, T. *et al. Cell* **22**, 767–775 (1980).
17. Czernilofsky, A. P. *et al. Nature* **287**, 198–203 (1980).
18. Itohara, S., Hirata, K., Inoue, M., Hatsuoka, M. & Sato, A. *Gann* **69**, 825–830 (1978).
19. Yoshida, M., Kawai, S. & Toyoshima, K. *J. Virol.* **38**, 430–437 (1981).
20. Hirt, N. *J. molec. Biol.* **26**, 365–369 (1967).
21. Zasloff, M., Ginder, G. P. & Felsenfeld, G. *Nucleic Acids Res.* **5**, 1139–1152 (1978).
22. Williams, B. G. & Blattner, F. R. *J. Virol.* **29**, 555–575 (1979).
23. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499–560 (1980).
24. Collet, M. S., Erikson, E., Purchio, A. F., Brugge, J. S. & Erikson, R. L. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3159–3163 (1979).
25. Hunter, T. & Sefton, B. M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1311–1315 (1980).
26. Neil, J. C., Ghysdael, J., Vogt, P. K. & Smart, J. E. *Nature* **291**, 675–677 (1981).
27. Beveren, C. V. *et al. Nature* **289**, 258–262 (1981).
28. Ellis, R. W. *et al. Nature* **292**, 506–511 (1981).
29. Levinson, A. D., Courtneidge, S. A. & Bishop, J. M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1624–1628 (1981).
30. Courtneidge, S. A., Levinson, S. A. & Bishop, J. M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3783–3787 (1980).
31. Czernilofsky, A. P. *et al. Nucleic Acids Res.* **8**, 2967–2984 (1980).
32. Swanstrom, R., Delorbe, W. J., Bishop, J. M. & Varmus, H. E. *Proc. natn. Acad. Sci. U.S.A.* **78**, 124–128 (1981).

LETTERS TO NATURE

Magnetic fields, convection and solar luminosity variability

W. C. Livingston

Kitt Peak National Observatory, Tucson, Arizona 85726, USA

Model calculations universally indicate that heat transport from the solar interior is almost entirely by convection in the outer envelope. Spiegel and Weiss[1] have recently discussed how magnetic fields can intrude into this envelope to affect convection and thus introduce transients into the Sun's luminosity. I demonstrate here from Fourier transform spectrometer observations of Fraunhofer line asymmetry that granular convection is retarded in the presence of surface magnetism. I then present full disk observations which suggest a lessening of convection over the past 5 yr. As this time interval coincides with the rise from a minimum to a maximum of solar activity, with the presumed injection of new magnetic flux from the interior, the resolved disk and full disk observations are consistent. Finally I note the continued temporal decrease in the spectroscopic temperature of the low photosphere and consider how a change of convection and the lowering temperature may be related.

The convective process is plainly manifest at the Sun's surface by the granulation pattern. Rising bright granules, transporting excess energy which is then radiated into space, are surrounded by the cooler intergranule gas which is falling. This causes a velocity–brightness correlation which in turn gives rise to an asymmetry in Fraunhofer lines observed near disk centre[2,3].

Spectral line asymmetries are traditionally represented by plotting the shifts of the bisectors of the line profile (as in Fig. 1). These bisectors have a characteristic 'C' shape, the middle of which is blueshifted in wavelength relative to the laboratory value. The upper end of the C is less blueshifted, as increased density low in the photosphere, coupled with the condition of flow continuity, results in lower convective velocities. The lower end is also less blueshifted, because of the cessation of upward velocity at the top of the granulation layer as represented by the line core. Detailed calculations by Nordlund[4] provide a quantitative basis for the above hypothesis.

A Fourier transform spectrometer (FTS) is preferred for bisector observations; it has a symmetric response profile, is capable of giving absolute wavelengths, records many lines simultaneously and is easily fed with unfocused (integrated) sunlight[5]. In a day-long sequence of full disk 1-m FTS measurements, mid C values typically display a wavelength variance of ~0.1 mÅ, corresponding to a velocity noise of 5 m s^{-1}. For historical reasons, some of the data presented here were obtained with a 13.5-m grating spectrometer. The greatest problem in this spectrograph is thermal drift which introduces a time-dependent asymmetric instrumental profile. Velocity noise is also about 10 times worse than with the FTS, but by averaging many records the noise in the data becomes tolerable.

Figure 1 shows line bisectors for a region of high magnetic flux compared with a non-magnetic area nearby. The spatial resolution is purposely low, about 1 × 2 arc min, to average over local conditions. This record is typical of some 50 regions

studied. In all cases near the disk centre the magnetic region line bisector has a smaller curvature and is less displaced to the blue than are bisectors for quiet regions. Although Fe 5,250.6 Å is a magnetically sensitive line (Landé $g = 1.5$), the $g = 0$ line Fe 5,576.1 Å gives the same result, demonstrating that we are not troubled by an instrumental artefact arising from telescope polarization, for instance, or other conceivable effects of Zeeman splitting. Dravins *et al.*[3] also find that line asymmetry is independent of *g*.

Figure 2*a* compares the full disk Fe 5,250.6 mean bisector for 1976–77 with the bisector for 1980–81. These are observations made with the 13.5-m spectrometer, the records having the worst thermal drift being eliminated, and 21 days of data (a total of 14 h) averaged together for each epoch. The error bars in wavelength are derived from the dispersion of the day averages. Errors in intensity are negligible. The maximum distance between the bisectors relative to the wavelength error occurs around an intensity level 20% down from the continuum. Here the formal error is 0.25 mÅ (12 m s^{-1}) while the separation is 1.5 mÅ (75 m s^{-1}).

Since May 1980 we have obtained full disk FTS bisector observations at 4–5-month intervals. As mentioned above, we have greater confidence in the FTS data than in those obtained with the spectrograph. For the FTS the error in wavelength is formally ~0.04 mÅ (2 m s^{-1}) based on six observations. The trend towards diminishing convection continues and is evident in the FTS bisector (Fig. 2*b*) even though solar maximum seems to have occurred in November 1979. Perhaps much of the flux which emerged at that time, largely concentrated in sunspots, is now slowly being distributed polewards to affect a growing fraction of the solar disk[6].

These bisector studies are an outgrowth of our line-strength monitoring programme which began in 1975 and which has been pursued vigorously. The cooling trend indicated early on

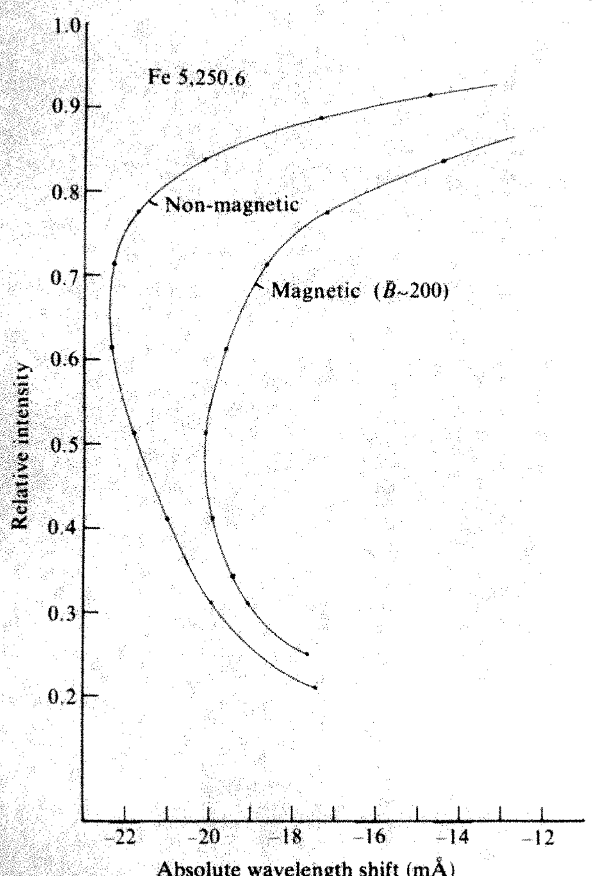


Fig. 1 Typical FTS line bisectors for adjacent magnetic and non-magnetic regions, observed at a spatial resolution of 1 × 2 arc min. The large, net blueshift arises from solar rotation. Note the non-magnetic C has larger amplitude and is more blueshifted than the magnetic C.

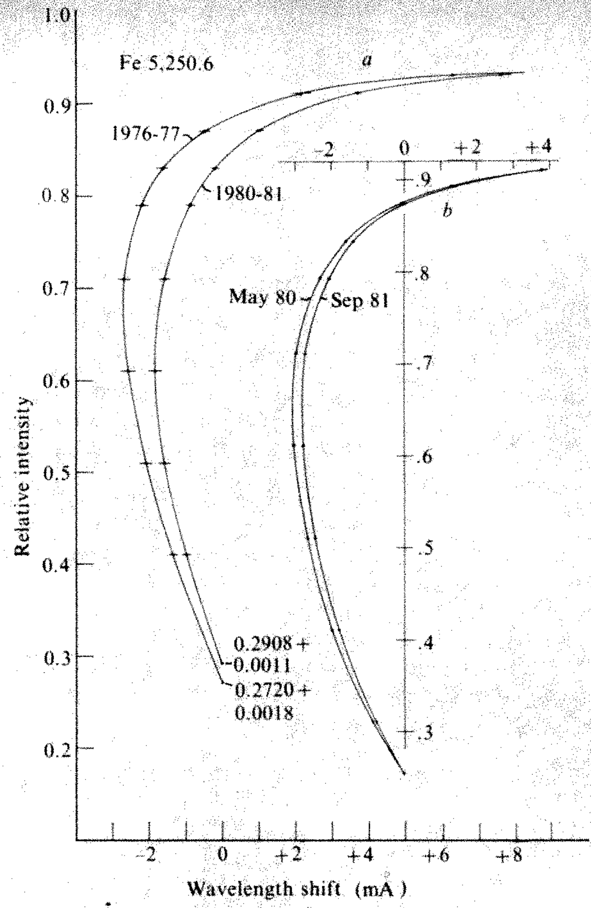


Fig. 2 *a*, Full disk spectrograph bisectors for solar minimum and maximum compared. The implication is decreased convection. *b*, As *a* except for a recent 16-month interval and using single day FTS data. In this case the formal errors are comparable with the size of the plotted points.

by a weakening of C 5,380 Å continues unabated. It was prematurely reported[7] that Fe 5,379 Å, which has the opposite temperature response to C 5,380 Å, was synchronously strengthening, as would have been expected. Instead, 4 yr of additional data show it unequivocally to be weakening also. This apparent inconsistency was resolved by Holweger[8] who used the fact that the two lines do not originate in the same atmospheric levels. He proposes that the low photosphere is temporally cooling, the upper photosphere is heating, and the result is a flattening of the temperature gradient. The quantitative implications of this change of temperature gradient on solar irradiance have not yet been determined.

I conclude that Fraunhofer lines in the Sun's integrated light spectrum display asymmetries which arise from convection. Observations resolving magnetic regions indicate that the amplitudes of the line asymmetries are smaller than in quiet regions. I have now observed a secular diminishing of the line asymmetry in full disk data from 1976 to 1981, corresponding to a period of increasing magnetic activity. Thus increasing surface magnetic flux (perhaps related to the phase of the activity cycle) is slightly retarding granular convection. Finally, this retarded convection may be the root cause of our observed line weakening and deduced solar cooling.

I thank Dr C. A. Lindsey for helpful suggestions.

Received 5 February; accepted 23 March 1982.

1. Spiegel, E. A. & Weiss, N. O. *Nature* **287**, 616–617 (1980).
2. Adam, M. G., Ibbetson, P. A. & Petford, A. D. *Mon Not. R. astr. Soc.* **177**, 687–708 (1976).
3. Dravins, D., Lindegren, L. & Nordlund, Å. *Astr. Astrophys.* **96**, 345–364 (1981).
4. Nordlund, Å. *Astr. Astrophysics* (in the press).
5. Brault, J. W. *Ossni Mem. Oss. astrofis. Arcetri* **106**, 33–50 (1979).
6. Howard, R. & LaBonte, B. J. *Solar Phys.* **74**, 131–145 (1981).
7. Livingston, W. C. *Nature* **272**, 340–341 (1978).
8. Livingston, W. & Holweger, H. *Astrophys. J.* **252**, 375–385 (1982).

Constraints on the Moon's origin from the partitioning behaviour of tungsten

Horton E. Newsom* & Michael J. Drake†

Lunar and Planetary Laboratory, and Departments of *Geosciences and †Planetary Sciences, The University of Arizona, Tucson, Arizona 85721, USA

The Moon, Earth and eucritic meteorites display W/La ratios which correspond to depletions of W relative to chondrites by factors of ~19 (ref. 1). This depletion can be attributed to metal/silicate fractionation. Ordinary chondrites and Cl chondrites with widely varying metal contents and oxidation states have similar W/La ratios, suggesting that nebular fractionations were unimportant. Therefore, assuming that both W and La are refractory and incompatible, the depletions of W relative to La in the Earth, Moon and eucrites are presumably due to separation of metal from silicate during planetary-scale igneous events. For the Moon, Rammensee and Wänke[2] concluded that any geophysically reasonable metallic lunar core would be too small to explain the large lunar depletion of W. It was therefore suggested[2] that the Moon acquired its depletion of W in a non-lunar igneous event, supporting the hypothesis that the Moon formed by fission from the proto-Earth, the upper portion of which was depleted in W due to terrestrial core formation[3-5]. We show here that this conclusion is implicity based on a special case, namely equilibrium between metal and silicate phases with total melting of silicates. In the geologically more general case where metal fractionates from silicate at relatively low degrees of partial melting, the incompatible nature of W may negate this conclusion. A geophysically plausible metallic core can account for the depletion of W in lunar surface and, presumably, lunar mantle rocks. The low W/La ratio in the Moon cannot be used as unconditional evidence for a terrestrial origin of the Moon.

We first consider the initial concentrations of W and La in the Moon or in precursor lunar material. The presence of correlated pairs of refractory lithophile elements (for example, Ba/La) with Cl ratios in the Earth, Moon and eucrites argues strongly for a chondritic W/La ratio in the Moon or precursor lunar material[6]. The abundances of W and La in lunar rocks and chondritic meteorites are plotted in Fig. 1. The magnitude of the lunar W depletion relative to Cl and ordinary chondrites was calculated from the data in Fig. 1 and is listed in Table 1. The range of depletion factors for the Moon relative to the chondrites is 19 to 29, but there are large uncertainties in these values due to the uncertainties in both the lunar and chondrite W/La ratios.

We next consider the amount of metal that may be present in the Moon. Geophysical evidence for the metal content of the Moon is inconclusive. The tightest constraint on the size of a lunar core, by Hood *et al.*[7], is a maximum radius ≤ 360 km, corresponding to 2% metal. Other investigations[8,9] find indirect

Table 1 W/La ratios in chondrites, and corresponding depletion factors for the Moon

Chondrite type	W/La ratios	Depletion factor (a) Moon/chondrite
Cl	0.36 ± 0.04	19 ± 7
H	0.55 ± 0.06	29 ± 11
L	0.45 ± 0.11	24 ± 10
LL	0.40 ± 0.09	21 ± 9

Lunar W/La ratio $= 0.019 \pm 0.007$, data from Fig. 1.

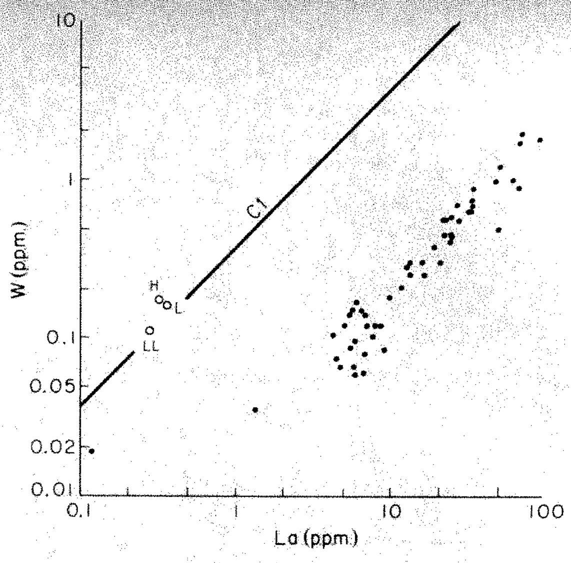


Fig. 1 Concentrations of W and La in lunar rocks, average values for three classes of chondrites, and the Cl chondrite W/La ratio[17-20].

indications of a metal core with a radius of 200 to 500 km. A 500-km radius core corresponds to ~5.5% metal.

With information on the possible magnitudes of the lunar W depletion and the possible lunar metal content, we now evaluate the parameters which control the partitioning of W between metal and silicate in the Moon. The most important parameters are the degree of partial melting, because of the incompatible nature of W, and the metal/(silicate liquid) partition coefficient for W. The partition coefficient depends on the oxidation state, which depends on the FeO content, the Ni content of the metal and the temperature of the Moon. The bulk composition and, hence, oxidation state of the Moon is uncertain. The effects of these parameters on the W/La ratio are illustrated in Figs 2 and 3 and discussed below.

The behaviour of W at low degrees of partial melting is particularly important. Tungsten is an incompatible element, with siderophile affinities in lunar conditions of low oxygen fugacity. Therefore, W prefers metal to silicate melt and prefers silicate melt to silicate solid. At low degrees of partial melting the concentration of W in the silicate melt and in the metal is very high because W is substantially excluded from solid silicates. Partitioning of W between solid metal and liquid metal is unimportant because the partition coefficient is close to unity[10]. Figure 2 shows that, at low degrees of partial melting, significant depletion of W relative to La can occur even if the metal/(silicate liquid) partition coefficient D(W) is relatively small. This is because the bulk metal/(silicate liquid + silicate solid) partition coefficient (bulk $D^{M/\Sigma S}$) is much larger than the D(W) at low degrees of partial melting.

The effective degree of partial melting in the Moon during metal/silicate fractionation is important to determine because of its large effect on the bulk $D^{M/\Sigma S}$. In their calculations Rammensee and Wänke[2] implicitly assumed that total melting of the silicates occurred, and that D(W) equals the bulk $D^{M/\Sigma S}$. In these conditions a very large D(W) is required if a geophysically plausible metallic core is responsible for the depletion of W in the Moon. It seems geologically more reasonable, however, that equilibrium between silicate and metal be confined to small degrees of partial melting, because metal will rapidly sink and the melt will rapidly rise before a large degree of partial melting can occur. As the metal starts sinking and aggregating, the decreasing time scale and increasing length scale for diffusion will rapidly terminate equilibration of the silicates with the metal[11]. A possible limit on the effective degree of partial melting is $F = 0.22$, where crystalline grains will no longer remain in contact with each other at the intergranular faces[12]. The investigation by Walker *et al.*[13] of the stability of

0028-0836/82/200210—03\$01.00

the terrestrial low-velocity zone may also be relevant. They found that for $F = 0.1$ the terrestrial low velocity zone could be largely drained of melt on a time scale of centuries, although this result is dependent on the scale of the melting system. Because of the large density difference between metal and silicate a value of $F < 0.1$ may be reasonable for the degree of partial melting during metal–silicate equilibration in the Moon, although the Moon may have continued to melt after effective isolation of the metal phase.

The other critical factor is D(W), the metal/(silicate liquid) partition coefficient for the Moon. D(W) depends strongly on the oxidation state of the Moon, which is related to the FeO content of the silicates and the Ni content of the metal in the Moon. The two scales plotted on the right ordinate of Fig. 2 show the relationship between D(W) and the FeO content of the silicates for both 6% Ni and 52% Ni in the metal: 6% Ni corresponds to the assumption that the Ni/Fe ratio in the metal is approximately the cosmic ratio; 52% Ni assumes a chondritic model for the relationship between Ni content of the metal phase and total metal content[14]. The actual Ni content of metal in the lunar interior probably falls within these limits. Model lunar compositions cover a wide range in FeO content[15] (2.3–13.9%), corresponding to a large range in D(W). Most of the recent model compositions tend towards higher values of FeO corresponding to lower values of D(W).

Figures 2 and 3 may be used to examine the conditions in which the depletion of W in the Moon could be explained by partitioning of W into metal within the Moon. The conditions discussed by Rammensee and Wänke[2] are a relatively high FeO content (13%), which corresponds to a low D(W) of 53, and total melting of the silicates ($F = 0.98$ in Fig. 3). For these

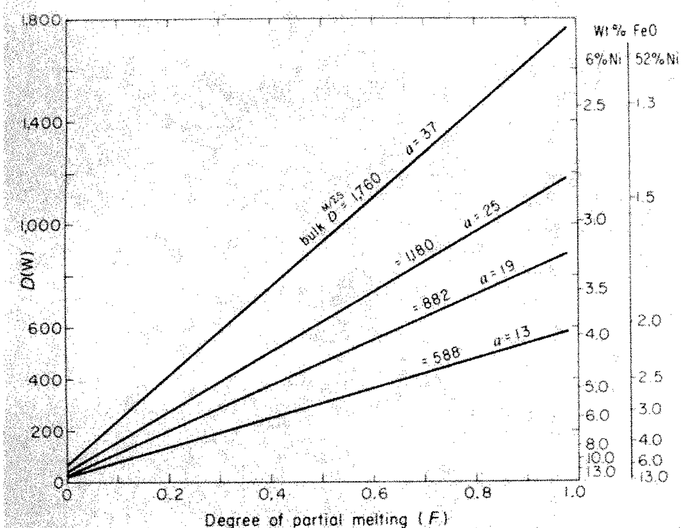


Fig. 2 The main constraints on the partitioning of W in the Moon constructed by calculating the bulk metal/(liquid silicate + solid silicate) partition coefficient (bulk $D^{M/\Sigma S}$), using an equilibrium partial melting model[21], for different degrees of partial melting (F) and different metal/silicate liquid partition coefficients, D(W). The solid silicate/liquid silicate partition coefficients used for both La and W are: olivine, 0.01; low-calcium pyroxene, 0.006; high-calcium pyroxene, 0.07; plagioclase, 0.14. Tungsten may be more incompatible than La, but no experimental mineral-melt data are known[22]. The model lunar silicate composition used was Hodges and Kushiro[15]. The diagonal lines represent the constant bulk $D^{M/\Sigma S}$ necessary to achieve a given depletion factor (a), assuming a metal content (x) of 2%. The relationship between bulk $D^{M/\Sigma S}$ and the depletion factor was calculated from the equation[2] $x = (a-1)/(\text{bulk } D^{M/\Sigma S} + a - 1)$. Breaks in slope occur when phases are exhausted, but the changes are small and are not shown. The two FeO scales plotted at the side were calculated by determining f_{O_2} (fugacity) for a given FeO content assuming 1,300 °C and ideal solution behaviour and 6% Ni or 52% Ni in the metal[23]. The corresponding D(W) was then calculated from the relation $\log D(\text{W}) = -\log f_{O_2} - 10.83$ determined experimentally by Rammensee and Wänke[2].

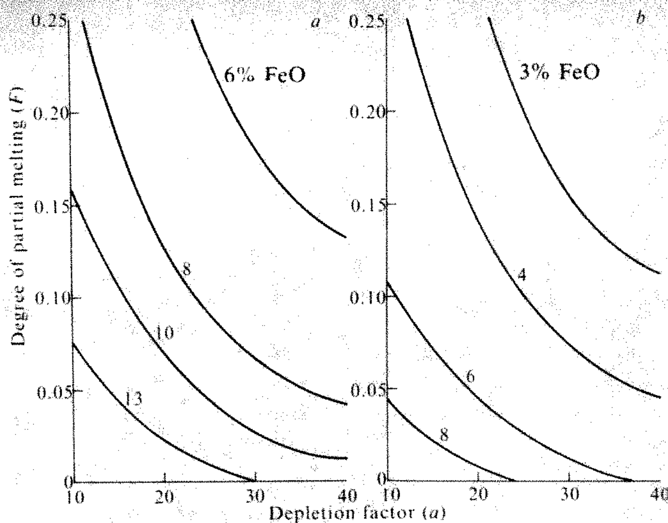


Fig. 3 a, b, The combination of per cent partial melting and FeO content required to achieve a reasonable depletion factor for the Moon (see Table 1): a, 2% metal containing 6% Ni; b, 2% metal containing 52% Ni. The plots were derived graphically from Fig. 2.

conditions Fig. 2 shows that a small metal content consistent with geophysical constraints could not be responsible for the observed depletion of W. We have argued, however, that equilibrium of silicate phases with metal at high degrees of partial melting is geologically unlikely. The situation at low degrees of partial melting is illustrated in a different way in Fig. 3a and b, which are derived directly from Fig. 2 for 6% Ni and for 52% Ni in the metal respectively. Figure 3a shows for the assumption of 6% Ni in the metal that reasonable lunar depletion factors (Table 1) can be achieved for lunar FeO contents as great as 13%. Figure 3b shows for the assumption of 52% Ni in the metal that reasonable depletion factors can be achieved only if the lunar FeO content is <6–8%. The actual ranges of values of D(W) and F which are consistent with the lunar depletion of W are uncertain due to the wide range in possible lunar depletion values (Table 1) and the uncertainty in the FeO content of mafic silicates and Ni content of metal in the Moon. The only situation where the low W/La ratio in the Moon could not be explained by a 2% metal core is the combination of a high lunar FeO content with a very high Ni content in the metal. Even this constraint would be greatly relaxed if a core larger than 2% metal exists[8,9].

Of less importance, but of significance, is the variation in the modal composition of the Moon for different lunar bulk compositions. Compositions with low FeO contents tend to have high normative plagioclase contents. A large amount of plagioclase reduces the bulk $D^{M/\Sigma S}$ at low degrees of partial melting, because plagioclase may have a higher solid silicate/liquid silicate partition coefficient for W than does olivine or pyroxene.

A major question regarding depletion of W during partial melting is explaining the very homogeneous W/La ratios in lunar samples. Figure 3 shows the large sensitivity of the depletion factors to small changes in the degree of partial melting. Several factors may help explain this problem. A closer look at the actual mechanism of metal segregation is needed. Assuming that the metal grains are not all the same size, the partitioning of W and the settling of metal grains probably occur over a significant range in the degree of partial melting. This suggests that an important control on the depletion of W is the size–frequency distribution of metal. A relatively constant size–frequency distribution and metal content of accreting material may lead to a relatively constant W/La ratio. Other mechanisms following core formation could also lead to homogenization. These include convection and the possible formation of a lunar magma ocean.

We conclude from the above analysis that a metal content consistent with present geophysical constraints on the size of a

metallic lunar core could account for the depletion of W observed in lunar rocks. The low W/La ratio of the Moon cannot be used as evidence for the formation of the Moon from the Earth's mantle by fission following terrestrial core formation. The approximate coincidence of W/La ratios in the Earth, Moon and eucrite parent body may only reflect the operation of the same depletion mechanism in similar conditions on several parent bodies[16].

We thank W. Rammensee for a helpful review. This work was supported by NASA grant NSG 7576.

Received 7 October 1981; accepted 23 March 1982.

1. Wänke, H. *et al. Proc. 4th lunar Sci. Conf.* 1461–1481 (1973).
2. Rammensee, W. & Wänke, H. *Proc. 8th lunar Sci. Conf.* 399–409 (1977).
3. Ringwood, A. E. *Origin of the Earth and Moon* (Springer, Berlin, 1979).
4. Delano, J. W. & Ringwood, A. E. *Proc. 9th lunar planet. Sci. Conf.* 111–159 (1978).
5. Wänke, H., Dreibus, G. & Palme, H. *Proc. 9th lunar Sci. Conf.* 83–110 (1978).
6. Dreibus, G., Kruse, H., Spettel, B. & Wänke, H. *Proc. 8th lunar Sci. Conf.* 211–227 (1977).
7. Hood, L. L., Sonett, C. P., Herbert, F., Cantrell, S. & Smith, L. *Lunar planet. Sci.* **12**, 457–459 (1981).
8. Russell, C. T., Coleman Jr, P. J. & Goldstein, B. E. *Lunar planet. Sci.* **12**, 914 (1981).
9. Stevenson, D. J. & Yoder, C. F. *Lunar planet. Sci.* **12**, 1043–1045 (1981).
10. Jones, J. & Drake, M. J. *Lunar planet. Sci.* **13**, 369–370 (1982).
11. Stevenson, D. J. *Science* **214**, 611–619 (1981).
12. Stocker, R. L. & Gordon, R. B. *J. geophys. Res.* **80**, 4828–4836 (1975).
13. Walker, D., Stolper, E. M. & Hays, J. F. *J. geophys. Res.* **83**, 6005–6013 (1978).
14. Anders, E. *Space Sci. Rev.* **3**, 583–714 (1964).
15. Warren, P. H. & Wasson, J. T. *Proc. 10th lunar Sci. Conf.* 2051–2083 (1979).
16. Newsom, H. E. & Drake, M. J. *Geochim. cosmochim. Acta* (submitted).
17. Rambaldi, E. R., Wänke, H. & Larimer, J. W. *Proc. 10th lunar planet. Sci. Conf.* 997–1010 (1979).
18. Palme, H. *et al. Proc. 9th lunar planet. Sci. Conf.* 25–57 (1978).
19. Mason, B. *U.S. geol. Surv. Prof. Pap.* 440-B-1 (1979).
20. Palme, H., Suess, H. E. & Zeh, H. D. in *Landolt-Börnstein* Group 6, Vol. 2, Part a, 252–257 (Springer, Berlin, 1981).
21. Consolmagno, G. J. & Drake, M. J. *Geochim. cosmochim. Acta* **41**, 1271–1282 (1977).
22. Palme, H. & Rammensee, W. *Proc. 12th lunar planet. Sci. Conf.* 949–964 (1981).
23. Myers, J. *EOS* **62**, 1070 (1981).

High-resolution coherent Raman spectroscopy of matrix-isolated carbon monoxide

I. R. Beattie, T. R. Gilson, S. N. Jenny & S. J. Williams

Department of Chemistry, The University, Southampton SO9 5NH, UK

IR spectroscopy of species isolated at low concentrations in inert matrices at cryogenic temperatures is an important physical technique in many branches of science. Raman spectra of such systems are much more difficult to obtain even at a resolution of only 1 cm^{-1}. Coherent Raman spectroscopy with an attainable resolution of at least 0.001 cm^{-1} is capable of rivalling diode laser or Fourier transform IR spectrometers. We report here the first coherent Raman spectrum of a matrix-isolated species. The system chosen is carbon monoxide in nitrogen (15K) at 0.13% concentration with a resolution of 0.13 cm^{-1}.

Although coherent Raman spectroscopy is experimentally demanding compared with spontaneous Raman spectroscopy it can enable experiments to be carried out that would otherwise be impossible. Perhaps the most spectacular results are those of Owyoung and co-workers[1] who used Raman loss spectroscopy (inverse Raman) to obtain spectra of $^{13}CH_4$ at 1.3 torr pressure. The resolution was Dopper limited at 0.009 cm^{-1} resolution. A more familiar coherent Raman experiment is coherent anti-Stokes Raman spectroscopy (CARS) which is experimentally easier to set up (and less expensive) than a sophisticated Raman gain/loss experiment.

The selection rules for both Raman gain/loss spectroscopy and CARS are the same as for those of the spontaneous Raman effect[2]. However there are important differences between the two coherent phenomena. Raman gain/loss spectroscopy is a two-photon process whereas CARS is a four-photon parametric process. The CARS intensity in phase-matched conditions is given by[2]:

$$I_3 = \left(\frac{8\pi^2 c^2 \omega_s}{h\omega_2^4}\right)^2 I_1^2 I_2 L^2 \sum_0^J \Delta N_J^2 \left(\frac{d\sigma}{d\Omega}\right)^2 \left[\frac{1}{2\Delta\omega_J - i\Gamma_J}\right]^2$$

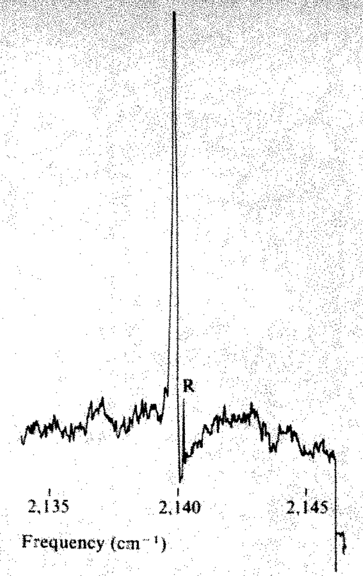


Fig. 1 The CARS spectrum of carbon monoxide in a nitrogen matrix at a concentration of 0.13%. (R = etalon reset.)

where ω_1, ω_2 and ω_3 and I_1, I_2 and I_3 are the pump (laser), Stokes (laser) and anti-Stokes (signal) frequencies and intensities respectively; ΔN is the population difference between ground and excited states; L is the interaction length; $d\sigma/d\Omega$ is the Raman scattering cross-section; $\Delta\omega_J$ is the Raman shift for the Jth resonance; and Γ a convolution of natural and laser line widths. Note the squared dependence on ΔN, concentration, interaction length and $1/\Gamma_J$.

Consideration of the above factors led us to attempt to obtain CARS spectra of matrix-isolated species at cryogenic temperatures. Matrix isolation spectroscopy involves the trapping of a molecule (atom or fragment) in an inert substrate at low concentration so that the molecule is well isolated from other molecules. Conventional matrices are the inert gases and nitrogen. As guest molecules do not normally rotate in the host matrix, the lines will be sharp (frequently <1 cm^{-1}). Furthermore, at the isolation temperatures (of the order of 10K), the molecules will be in their ground vibrational states. Thus ΔN and $1/\Gamma_J$ are both maximized. For these experiments we chose nitrogen as the matrix gas because: (1) it gave us a strong signal from the substrate for initial lining up; (2) it has the advantage that the $^{14}N^{15}N$ (0.7 mol% concentration) peak may be used as a test for alignment and sensitivity. Carbon monoxide was chosen as the species to be isolated because not only has it been extensively studied, but also it has a vibrational frequency close to that of nitrogen.

Figure 1 shows the first coherent Raman spectrum of a matrix-isolated species. The concentration of the carbon monoxide is 0.13% and the resolution of the experiment is of the order of 0.13 cm^{-1}, which is an order of magnitude better than that obtained in a conventional matrix isolation Raman experiment. A further order of magnitude improvement in resolution could be obtained without significantly altering the nature of this experiment.

The spectrum was recorded using a carbon monoxide–nitrogen premix pulsed on to a calcium fluoride plate held in a Displex closed-cycle cooler operating in the region of 15K. The CARS system used a J K Lasers Q-switched Nd/YAG laser as pump, operating at 16.7 Hz repetition rate, with a 10 ns pulse length. The probe laser was a Chromatix CMX4 flash lamp pumped scanning dye laser amplified to ~100 kW peak power. For colourless matrices we find a laser damage threshold at a few hundred kilowatts peak power for a focused beam diameter of ~10^{-2} cm. The spectra were recorded by scanning the CMX4 synchronously with (but in the opposite sense to) a Cary 81 double monochromator used as a high quality filter. The signal was detected by conventional means using a photo-

0028-0836/82/200212—02$01.00

multiplier and boxcar detection. The frequency of the carbon monoxide transition was calibrated by running IR spectra on the same matrices, as the CARS spectra in these experiments were only recorded to an accuracy of $\pm 1\ cm^{-1}$.

Attempts to carry out studies in argon led to poor-quality spectra, which were only obtained at concentrations of the order of 1%. This is clearly due to line broadening coupled with interference between the resonant signal and the non-resonant background, leading to the observation of derivative type spectra. (There are several techniques available for minimizing this effect but they all lead to a reduction in signal.) In nitrogen the rotation of the carbon monoxide (which is thought to account for the line broadening in argon[3]) is presumably suppressed, leading to a narrow line.

Note that with Raman gain/loss spectroscopy a conventional Raman spectrum is obtained without the complications outlined above for CARS. There is also the possibility of a further order of magnitude improvement in resolution (leading to $0.001\ cm^{-1}$ as a reasonable limit).

Received 31 December 1981; accepted 30 March 1982.

1. McDowell, R. S., Patterson, C. W. & Owyoung, A. *J. chem. Phys.* **72,** 1071–1076 (1980).
2. Tolles, W. M., Nibler, J. W., McDonald, J. R. & Harvey, A. B. *Appl. Spec.* **31,** 253–272 (1977).
3. Dubost, H., Lecuyer, A. & Charneau, R. *Chem. phys. Lett.* **66,** 191–194 (1979).

Rapid localized glacio-isostatic uplift at Glen Roy, Scotland

J. B. Sissons & R. Cornish

Department of Geography, High School Yards, Edinburgh EH1 1NR, UK

Many studies involving measurement of the altitudes of fossil marine and lacustrine shorelines have demonstrated glacio-isostatic tilting, particularly in North America, Fennoscandia and the British Isles. Such studies are normally illustrated by height–distance diagrams in which fossil shorelines are portrayed as straight lines. In some studies the shoreline-relation diagram is used to analyse and illustrate the altitudinal data[1–3]: before constructing such a diagram it is assumed that a simple pattern of glacio-isostatic uplift has prevailed. Warping of shorelines, as opposed to uniform tilting, has been inferred in North America[4–6]. On the other hand, Härme[7] envisaged block movements in Fennoscandia. Sauramo[8] also proposed dislocations, but his interpretation was rejected by Hyyppä[9]. More recently the concept of the Scandinavian shield as a stable area has been dismissed[10,11]. Here we provide evidence from Scotland for shoreline dislocation by differential movement of blocks of the Earth's crust.

The study of former crustal movements from fossil shorelines requires accurate, very closely spaced measurement, which has rarely been attempted outside Scotland. The results of such a study in the Forth Valley in central Scotland are shown in Fig. 1*a*. The upper shoreline, formed at 6,800 yr BP, is a surface feature while the lower shoreline, formed at 9,600 yr BP, is a buried feature[12]. Both shorelines have corresponding sloping and essentially horizontal portions. Two parts of this area have been glacio-isostatically uplifted without significant tilting in the past 9,600 yr, while the remainder has been tilted. Furthermore, the older shoreline has two clear displacements, one of 1 m at the limit of the Loch Lomond (Younger Dryas) Advance, and one of nearly 1.5 m corresponding with a major fault: these displacements had ended by 6,800 yr BP (Fig. 1*a*).

In Glen Roy and vicinity in the Scottish Highlands, ice-dammed lakes existed during the Loch Lomond Stadial. We have recently levelled their fossil shorelines at 1,760 points. Measurements were made at the break of slope representing the former shoreline (junction of 'parallel road' and backing cliff) and were spaced, where possible, at 25–30-m intervals. Traverses were normally 4–5 km long and the average closing error was 0.011 m. Where peat was present probes 1–3 m apart were aligned perpendicular to the shoreline. This work is more detailed and accurate than any previously undertaken on shorelines of former lakes.

We find that the shorelines are in part horizontal and in part tilted while, occasionally, there are displacements of <2 m. Despite these variations each shoreline has a limited altitudinal range. Using the means for groups of measured points and, where linear regression analysis was possible, the calculated values for the ends of tilted lengths of shoreline, the highest shoreline of Glen Roy varies between 349.5 and 351.9 m, the middle shoreline between 324.5 and 326.8 m, and the lowest one between 260.1 and 262.4 m, ranges of 2.4, 2.3 and 2.3 m respectively. There is a major exception to these generalizations which we consider here.

On the north-west side of upper Glen Roy a large landslide has destroyed all three 'parallel roads' (Fig. 2). Immediately south-west of the landslide all three shorelines rise above their normal altitude for this general area by ~3 m: the highest shoreline attains 354.1 m, the middle one 327.7 m and the lowest one 263.9 m. The 'roads' are here cut into a steep slope (25–30°) composed of metamorphic rocks of the Eilde Flags Formation (Moinian). Drift cover is thin and is mainly weathered bedrock. South-west from the landslide linear depressions roughly parallel with the contours mark the upper edges of slipped masses of bedrock (Fig. 2): they appear to be clear evidence of downhill movement but occur even where levelling proves the 'roads' to have been uplifted most.

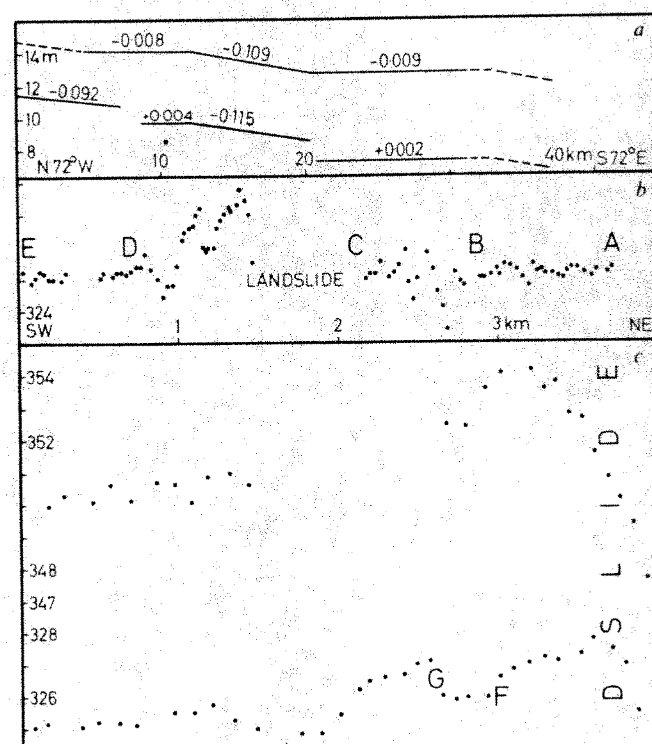


Fig. 1 Height–distance diagrams of dislocated shorelines. *a*, Western Forth Valley, gradients (in m per km). *b*, The middle shoreline of Glen Roy on both sides of the major landslide. *c*, All three Glen Roy shorelines south-west of the major landslide.

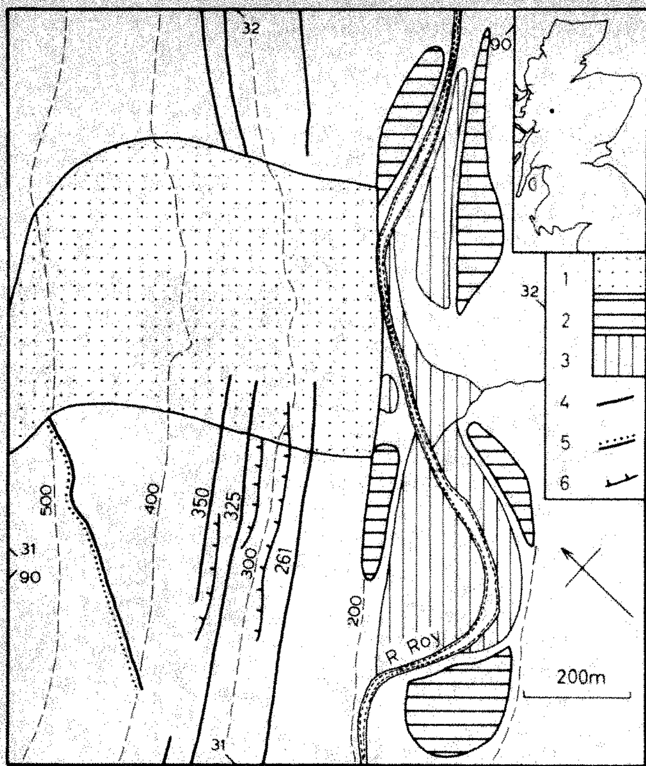


Fig. 2 The major landslide and other features. 1, Landslide; 2, part of the uniformly-sloping river terrace mentioned in the text; 3, other river terraces and floodplains; 4, 'parallel roads'; 5, inferred fault scarp; 6, upper edges of slipped masses of bedrock.

The middle shoreline, which is the best developed and sharply defined, is horizontal away from the major landslide: 16 levelled points in the south-west and 25 in the north-east (AB, DE, Fig. 1*b*) together average 325.1 m, the standard deviation being only 0.15 m. Between B and C the 'parallel road' is broken into several tilted portions: two of these rise to 0.7 m above shoreline altitude in the area of horizontality, indicating that BC is part of the uplifted area. South-west of the landslide the middle shoreline attains 2.6 m above its horizontal altitude. In the same locality the lowest shoreline attains 2.8 m above its altitude in the horizontal area and the highest shoreline 3.3 m above. All three 'roads' extend a short distance into the ground occupied by the major landslide, which has carried them down with it (Fig. 1*c*).

The point of maximal uplift on the highest shoreline lies upslope from the comparable point on the lowest shoreline. The culmination point of the middle shoreline is offset by 100 m: this difference is readily explained by the minor slips (for example, F and G, Fig. 1*c*) that have produced small but clear steps in the middle 'road'.

Localized uplift of the shorelines cannot be attributed to the landslide. Upward movement at the toe of a rotational landslide is well known: however, the Glen Roy feature is not a simple rotational slip and the toe is at the valley floor. Furthermore, the uplifted area extends hundreds of metres from the landslide. It is therefore most likely that crustal deformation occurred.

Pronounced uplift in a very small area suggests fracturing of the Earth's crust. Such fracturing is consistent with the presence of a linear bedrock ridge a few metres wide that slopes obliquely up the side of Glen Roy for a distance of 550 m (Fig. 2). Between the ridge and the valley side is a narrow depression that is up to 2 m deep despite being partly infilled by frost debris from the valley side and by peat. This feature is not a meltwater channel: the ridge is a clear positive feature standing above the adjacent ground (vertical aerial photographs reference O.S. 64:233, 032–033). Northwards the feature passes into the landslide. Southwards it begins to fade out before disappearing beneath frost debris that has moved downslope. Projected southwards the feature coincides exactly with the point where the middle shoreline begins to rise above its normal altitude (D, Fig. 1*b*). Part of the top 'road' is missing in the critical area and the bottom 'road' is here too poorly preserved to determine a relationship.

The linear nature of the ridge and associated depression and the evidence for uplift imply a fault. No detailed geological map of the area exists. Anderson[13] mapped a straight fault 10 km to the NNE the projection of which coincides with the Glen Roy feature. The bend in the presumed fault corresponds with a small re-entrant in the hillside suggesting, along with the ridge, the presence of a reverse fault: this is consistent with crustal compression associated with glacio-isostatic uplift. Fault movement could have caused earthquakes, thus accounting for the large landslide.

The uplift clearly post-dates formation of the 'parallel roads'. The relative date of termination of uplift is also established. When the lake was emptied, probably very rapidly by *jökulhlaup*[14], the new land surface encountered by the River Roy in the locally uplifted area was undulating, with large gravel fans alternating with lower ground. Initially the Roy cut into the fans and deposited the gravel thus obtained between them, producing a smooth profile now represented by a river terrace that extends intermittently along the whole length of the 2-km long uplifted area. The edge of the terrace is 110–170 m distant from the lowest shoreline. The standard deviation from a linear regression line ($r = 0.997$) of 41 levelled points on the terrace is 0.45 m. Hence pronounced differential uplift in this area had ceased by the time the terrace was formed. It is likely that the River Roy established this smooth profile shortly after lake drainage for the terrace is entirely in loose material (gravel) and, farther up the glen, was subsequently dissected by the Roy into a complex sequence of lower terraces. The major climatic amelioration that terminated the Loch Lomond Stadial occurred while the Roy ice-dammed lake stood at the highest shoreline. Hence the pronounced localized uplift occurred at the beginning of the Holocene.

Although minor earthquakes occur in Scotland today no related fault scarp has ever been detected. The country is essentially stable compared with the pronounced former instability implied by the Glen Roy evidence we have given. The latter is therefore attributed to glacio-isostatic uplift.

Our Glen Roy evidence thus shows that glacio-isostatic recovery can include abrupt localized uplift in an area now relatively stable. It adds to the Forth evidence (Fig. 1*a*) demonstrating differential movement of blocks of the Earth's crust. Because in both areas the dislocations post-date the Loch Lomond Advance, when glaciers were of limited extent, we suspect that larger movements occurred earlier during ice-sheet decay. The probability clearly exists that in Fennoscandia and North America, where ice sheets were far larger than in the British Isles, there are shoreline dislocations of even greater magnitude.

The NERC financed this research. Dr D. G. Sutherland criticized the draft of the paper.

Received 17 November 1981; accepted 26 March 1982.

1. Tanner, V. *Bull. Commn. géol. Finl.* **88,** 1–594 (1930).
2. von Post, L. *Geol. För. Stockh. Förh.* **69,** 293–320 (1947).
3. Synge, F. M. & Stephens, N. *Trans. Inst. Br. Geogr.* **39,** 101–125 (1966).
4. Goldthwait, J. W. *J. Geol.* **16,** 459–476 (1908).
5. Leverett, F. & Taylor, F. B. *U.S. geol. Surv. Monogr.* **53,** 529 pp. (1915).
6. Broecker, W. S. *J. geophys. Res.* **71,** 4777–4783 (1966).
7. Härme, M. *Fennia* **89,** 29–31 (1964).
8. Sauramo, M. *Suomal. Tiedeakat. Toim.* III, **51** (1958).
9. Hyyppä, E. *Fennia* **89,** 37–48 (1964).
10. Lundqvist, J. & Lagerbäck, R. *Geol. För. Stockh. Förh.* **98,** 45–51 (1976).
11. Stephansson, O. & Carlsson, H. in *Earth Rheology, Isostasy and Eustasy* (ed. Mörner, N.-A.) 327–337 (Wiley, Chichester, 1980).
12. Sissons, J. B. *Trans. Inst. Br. Geogr.* **55,** 145–159 (1972).
13. Anderson, J. G. C. *Trans. R. Soc. Edinb.* **63,** 15–36 (1956).
14. Sissons, J. B. *J. geol. Soc. Lond.* **136,** 215–224 (1979).

Exotic terranes of western California

M. O. McWilliams

Geophysics Department, Stanford University, Stanford, California 94305, USA

D. G. Howell

US Geological Survey, Menlo Park, California 94025, USA

Numerous distinct geological terranes compose the North American Cordillera[1]; there may be as many as 50 terranes in California alone[2]. Critical to deciphering the history of Cordilleran tectonic assembly is an understanding of the displacement history of individual terranes. It is therefore important to know: (1) whether a terrane has undergone significant motion with respect to the stable craton (that is, whether it is allochthonous or exotic); (2) if so, when relative motion started and stopped; (3) from where an individual terrane originated; and (4) the nature of inter-terrane movements. We consider here the problem of determining whether the now-juxtaposed Salinian and Stanley Mountain terranes of California became amalgamated at or near their present position with respect to cratonic North America, or if they collided at a considerable distance from their present positions and were later accreted to North America as a composite package. The palaeomagnetic data that we present indicate that the latter was the case.

Terranes are fault-bounded geological entities of regional extent, each characterized by a geological history that is different from that of adjacent terranes. Such geological histories are determined from stratigraphical successions preserved within terranes, but in some cases this information is obscured owing to tectonic or sedimentological disruption or metamorphic overprinting. Composite terranes are produced by the amalgamation of two or more terranes. The southern Coast Ranges of California west of the San Andreas Fault consist of two composite tectonostratigraphical packages called the Salinian and Sur-Obispo terranes (Fig. 1). Basement rocks of the Salinian terrane are Middle and Upper Cretaceous granite plutons and older high-temperature metasedimentary rocks bounded on the west by the Sur-Nacimiento Fault. To the west lies the Sur-Obispo composite terrane, represented by the Stanley Mountain and San Simeon terranes. The Stanley Mountain terrane consists of Middle Jurassic ophiolite and an overlying thick sequence of Upper Jurassic to Upper Cretaceous forearc-basin strata. Structurally below the Stanley Mountain terrane is Franciscan mélange of the San Simeon terrane. The Stanley Mountain strata closely resemble, and are coeval with, the Great Valley sequence[3], which is typified by exposures along the west sides of the Sacramento and San Joaquin Valleys. Truncation by faults and unconformities disrupts the section in most places, but a nearly complete section of 4,000 m of Upper Jurassic and Cretaceous strata is exposed in the southeastern San Rafael Mountains[4]. These forearc-basin strata are demonstrably allochthonous with respect to Salinian basement, yet in late Cretaceous time the two terranes seemingly collided to form a new amalgamated terrane, as Campanian and younger sedimentary rocks of both terranes are petrographically identical and are characterized by non-marine horizons of locally derived granitic detritus[5,6] (Fig. 2). Several coherent packages of Upper Cretaceous arkosic strata in the San Simeon terrane may indicate continuity with the Salinian and Stanley Mountain terranes; however, equivocal sedimentary sequences which lap across these terranes are no older than early Tertiary.

Commencing in Campanian time with the paralic granite-grus lithofacies, the forearc took on a borderland-like configuration with detritus spilling into contrasting depositional settings on the newly juxtaposed terranes[6,7]. An episode of wrench faulting accompanied the suturing presumably along a now-distant margin of the Cordillera because Upper Cretaceous and Lower Tertiary strata contain numerous clasts of siliceous volcanic porphyries admixed with the granitic debris, yet such a volcanic provenance does not exist on either the Salinian or Stanley Mountain terranes.

Palaeomagnetic inclination data from Upper Cretaceous strata of the Salinian terrane indicate that deposition occurred as much as 2,500 km south of their present location (ref. 8; M. E. Williams, M. Schnapp and A. Cox, in preparation; D. E. Champion, personal communication). We report here new data which indicate a similar exotic origin for the Jurassic and Cretaceous rocks of the Stanley Mountain terrane.

To determine the palaeolatitude of the Stanley Mountain terrane in Jurassic time, samples of pillow basalt were collected at 10 sites spanning an 800 m section of ophiolite at Stanley Mountain (Figs 1, 2). Alternating field (a.f.) demagnetization of pilot specimens from each site yielded stable, well grouped magnetization directions at eight sites after removal of young secondary magnetizations of probable recent viscous origin; one example is shown in Fig. 3*a*. After partial demagnetization in peak alternating fields of 60–80 mT, the eight mean directions are well grouped and define an overall *in situ* mean direction $D = 053°$, $I = +54°$, $\alpha_{95} = 8°$. This direction is distinct from the present geomagnetic and axial dipole field directions for the sampling locality, as well as from any expected field direction for stable North America since early Mesozoic time[9]. A reasonable explanation for the *in situ* direction is that it represents a sampling of a primary Jurassic magnetization (TRM or high-temperature CRM) and should therefore be restored to palaeohorizontal attitude, thereby obtaining the palaeolatitude at which the rocks acquired their magnetization. If the magnetization is primary, the overall tectonically corrected mean direction is $D = 041°$, $I = -27°$, $\alpha_{95} = 8°$, corresponding to a palaeolatitude of $14 \pm 7°$. Structural attitudes of sedimentary rocks which immediately overlie the basalts were used as a palaeohorizontal reference indicator. If the age of the pillow basalt is 160 ± 5 Myr, as indicated by isotopic data from other Coast Range ophiolite samples[10], then a northerly or southerly

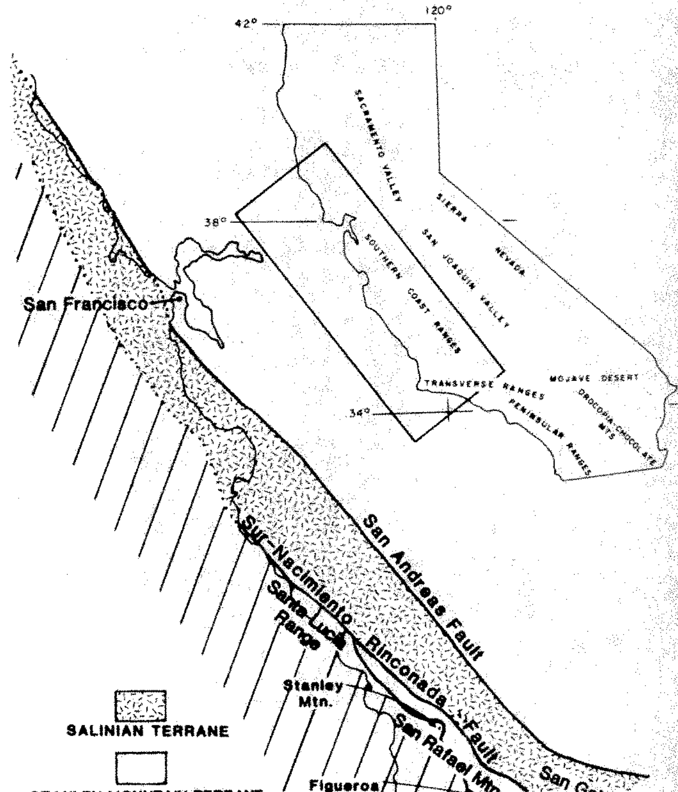


Fig. 1 Geographical sketch map of Stanley Mountain, San Simeon and Salinian terranes in western California.

0028-0836/82/200215—03$01.00

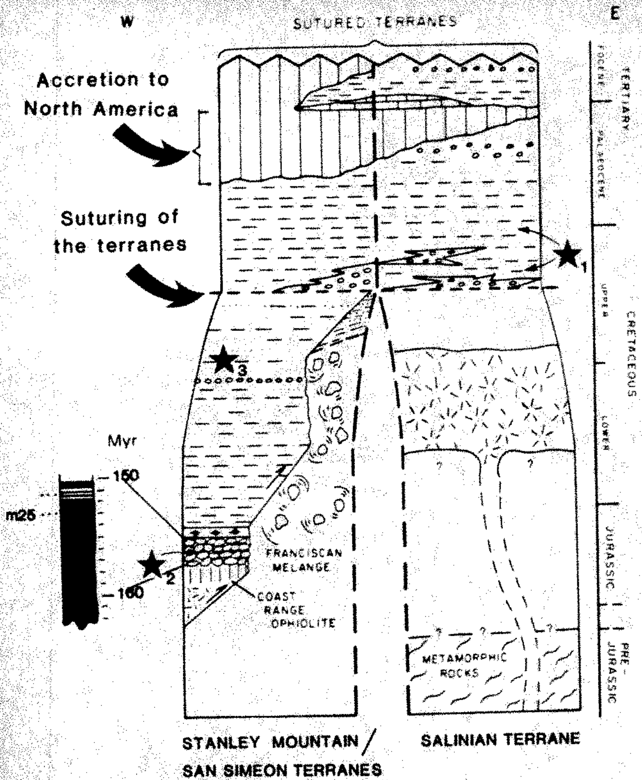


Fig. 2 Generalized stratigraphical columns for Salinian and Sur-Obispo composite terranes. Suturing of the two terranes was probably completed by the late Cretaceous, as equivalent strata lap onto both terranes after this time. ★, Stratigraphical positions of palaeomagnetic studies (*a*, D. E. Champion, personal communication; *b*, ref. 8 and M. E. Wilson, M. Schnapp and A. Cox, in preparation). The inset on the left is magnetic reversal time scale[11] (black = normal polarity).

palaeolatitude cannot be invoked, for the geomagnetic field was of mixed polarity at this time[11] (see reversal time scale inset, Fig. 2).

If the magnetization were of secondary origin and dated from Tertiary time, an exact tectonic correction would not be possible because the structural attitude at the time of remagnetization would be unknown. One possibility would be that the basalts acquired their magnetization after tectonic tilting and were subsequently rotated 50–60° clockwise about a vertical axis. At the locality where the samples were collected, the basalts form part of a steeply dipping homoclinal sequence, so that a fold test was not possible. There is no evidence, however, for Tertiary remagnetization by a thermal mechanism, for isotopic ages are not reset and reheating effects are not visible in hand specimen or thin section.

The sedimentary record indicates that the Stanley Mountain and Salinian terranes probably collided in the late Cretaceous. To determine the palaeolatitude of the Stanley Mountain terrane at approximately the time of collision, samples of Upper Cretaceous (Cenomanian–Santonian) flysch strata were collected at 10 sites in the San Rafael Mountains near Figueroa Mountain (Figs 1, 2). After a.f. demagnetization in peak fields up to 80 mT, a single stable component of magnetization was isolated (an example is shown in Fig. 3*b*). The overall *in situ* mean direction from these 10 sites is $D = 002°$, $I = +74°$, $\alpha_{95} = 10°$, which is significantly different from both the present geomagnetic field direction and the axial dipole field direction at the sampling locality, suggesting that it is not of recent origin. Also, this overall mean direction is significantly different from any geomagnetic field direction since early Mesozoic time derived from stable North America[9]. On tectonic correction to palaeohorizontal, the overall mean direction is $D = 023°$, $I = +12°$, $\alpha_{95} = 11°$. The minor decrease in overall precision accompanying this correction is not statistically significant and may be due to pre-magnetization slumping as noted in other palaeomagnetic studies of turbidites[8]. The slightly skewed distribution of site mean directions may be due to our inability to make exact tectonic corrections which compensate for this post-depositional slumping.

If the magnetization of these samples is primary and dates from the time of deposition or soon thereafter, the tectonically corrected inclination of $12 \pm 10°$ would correspond to a palaeolatitude of $6 \pm 5°$ (N or S). This range is slightly lower than the range of apparent palaeolatitudes from the Pigeon Point Formation on the Salinian terrane (ref. 8, and M. E. Wilson, M. Schnapp and A. Cox, in preparation), but if a Northern Hemisphere interpretation is invoked (Fig. 4), this might be expected, as the Pigeon Point strata are slightly younger (Campanian–Maestrictian). An alternative explanation for the palaeolatitude discrepancy could be that the suturing of the two terranes might not have been completed by this time.

The palaeomagnetic data shows that (1) the Salinian composite terrane lay in low ($21 \pm 7°$) palaeolatitudes[8] in Campanian–Maestrictian time, and occupied higher ($25 \pm 3°$) palaeolatitudes[10] in early Palaeocene time; (2) the Stanley Mountain terrane lay in slightly lower ($6° \pm 5°$) palaeolatitudes in Cenomanian–Santonian time, and occupied slightly higher ($14° \pm 7°$) palaeolatitudes at about 160 ± 5 Myr; either a north or south palaeolatitude is possible; (3) a simple displacement model (solid line in Fig. 4) would place the Sur-Obispo composite terrane at 14° S at 160 Myr, at 6° N at about 90 Myr and at 21° N at about 70 Myr, after colliding with the Salinian composite terrane; and (4) alternatively, the Stanley Mountain terrane could have occupied low northerly palaeolatitudes from Jurassic to Cretaceous time, beginning its major northward migration at about the time it collided with the Salinian terrane (dashed line in Fig. 4). Coupled with the geological arguments, these data indicate that the late Cretaceous collision of the two terranes occurred in low palaeolatitudes, far from the present California continental margin.

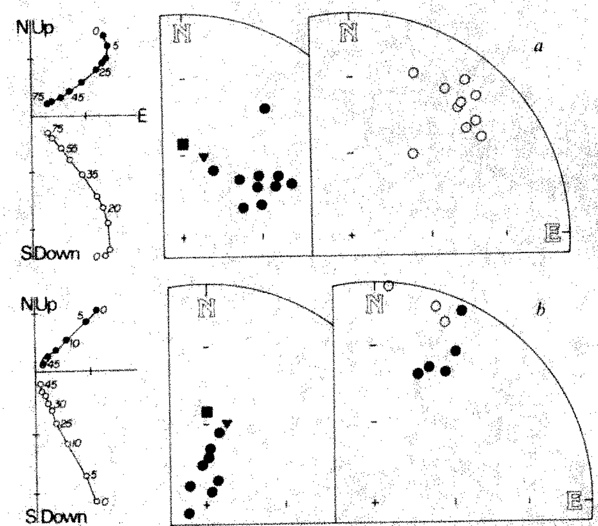


Fig. 3 *a*, Palaeomagnetic results from Stanley Mountain pillow basalts. Left: orthogonal vector projection of a.f. demagnetization experiment; circles represent projection of magnetization vector on the horizontal (●) or east–west vertical (○) plane; axis ticks 0.2×10^{-3} A m^{-1}. Numbers are peak a.f. treatment in units of 10^{-3} T. Centre: equal area projection of site mean magnetization directions after a.f. demagnetization. No tectonic correction is applied; each direction plotted is the average of six individually oriented samples. Solid symbols represent positive (downward) inclinations. ■ and ▼ are axial dipole field and present geomagnetic field directions respectively. Right: as in centre, but with magnetization directions restored to palaeohorizontal. Open symbols represent negative (upward) inclination. *b*, Palaeomagnetic results, Upper Cretaceous sedimentary rocks from Figueroa Mountain. Left: as above, axis ticks 0.1×10^{-3} A m^{-1}. Centre: site mean directions before tectonic correction; conventions as above. Right: site mean directions before tectonic correction; conventions as above.

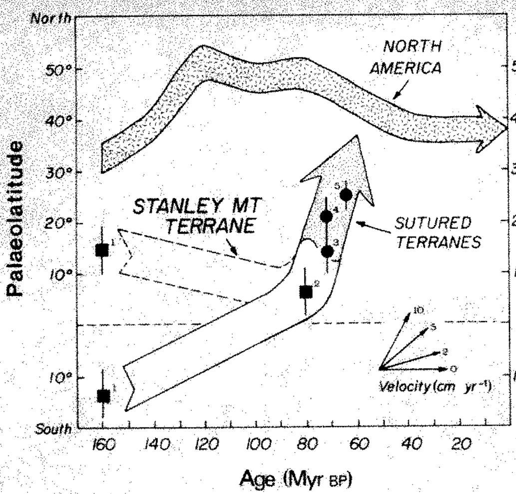


Fig. 4 Palaeolatitude of the Salinian and Stanley Mountain terranes as a function of time, compared with palaeolatitude of 'target area' on North America for present-day position, 35–39° north latitude. Stippled path represents the two terranes after suturing. The North American and Salinian/Stanley Mountain tracks intersect at ~50–55 Myr BP to allow for Eocene reconstruction of the Butano-Point of Rocks submarine fan system. ■, Data from Sur-Obispo terrane; ●, data from Salinian terrane. (1) and (2) are from this study; (4) from ref. 8; (3) from M. E. Wilson, M. Schnapp and A. Cox, in preparation; and (5) D. E. Champion, personal communication). Arrows in lower right give northward velocity in cm yr^{-1} for reference. The dashed and solid lines represent alternative palaeolatitudes for the Stanley Mountain terrane depending on the choice of magnetic polarity.

The classic sequence of events proposed for the evolution of the western Cordillera involves three general episodes[12-14]: (1) a late Jurassic to early Tertiary subduction regime with major deposition in an arc–trench gap configuration; (2) early Tertiary proto-San Andreas faulting and a borderland setting during an early to middle Tertiary episode of oblique subduction along a flattened Benioff–Wadati zone; and (3) continental margin transform faulting resulting in tectonic splintering and additional borderland development. If recent palaeomagnetic studies (ref. 8; M. E. Wilson, M. Schnapp and A. Cox, in preparation; D. E. Champion, personal communication), including our own, are accurate and basin analyses[4,6,7] correct, both the timing of events and nature of origin of the continental margin must be reinterpreted.

Analysis of the 1,500 m thick fluvio-deltaic sequence in the southeastern Stanley Mountain terrane indicates that the Stanley Mountain and Salinian terranes were probably attached by ~72 Myr BP. Thus the Stanley Mountain and Salinian terranes, both of which come from quite different parent terranes, must have travelled northwards together, despite the remarkable parallelism of Cretaceous depositional environments between the Stanley Mountain terrane and the Great Valley sequence of the Sacramento and San Joaquin Valleys. However, the Stanley Mountain/Salinian composite package was probably emplaced in California by the early Eocene, based on the inference that submarine fans sequences of this age[15] lap across these terranes and other terranes autochthonous to California.

The palaeomagnetic data presented here and elsewhere indicate a large northward component of relative plate motion along the western margin of North America during the Cretaceous and early Tertiary. Several late Mesozoic terranes presumably originated at the latitude of present-day Central America and moved northwards to the California area by late Palaeocene time. Plate motion reconstructions (ref. 16 and D. C. Engebretson, in preparation) demonstrate that Farallon–North America relative motion for most of the Cretaceous would have been largely orthogonal to a north-west trending margin of western North America. Kula plate motion relative to North America had a significant northward component at high latitudes, but owing to an Euler pole position in northwestern South America, Kula–North America relative motion in the Central American regions would also have been largely orthogonal (ref. 17 and D. C. Engebretson, in preparation).

Two possible ways to account for the northward shifting of terranes are immediately apparent[17]: (1) reconstruct the late Mesozoic Pacific basin plates including now vanished plate(s) whose configurations would account for the northward slip required by the palaeomagnetic data and for the episodic plutonism mapped all along the Cordillera, and/or (2) palinspastically restore Central America and the ancient Caribbean plate westwards to increase the distance from the Kula–North America Euler pole in late Mesozoic time, so that Kula plate motion could better account for the northward component of motion of the Stanley Mountain and Salinian terranes.

We thank Anne Demitrack, Brian Hausback, Bill Keller, Peter Lellis and Norman Sleep for assistance in sample collection, David Engebretson and Ben Page for editorial comments, and especially Michel Debiche and Anne Demitrack for many laboratory measurements. The research was supported by grant PRF-11405 from the Petroleum Research Fund of the American Chemical Society.

Received 21 January; accepted 9 March 1982.

1. Coney, P. J., Jones, D. L. & Monger, J. W. H. *Nature* **288**, 329–331 (1980).
2. Blake, M. C., Howell, D. G. & Jones, D. L. *U.S. geol. Surv. Open File Map Ser.* (in the press).
3. Page, B. M., in *The Geotectonic Development of California* (ed. Ernst, W. G.) (Prentice-Hall, Englewood Cliffs, New Jersey, 1981).
4. Howell, D. G. *et al.* in *Soc. Economic Paleontologists and Mineralogists, (Pacific Sec.), Pacific Coast Paleogeogr. Field Guide* Vol. 2 (eds Howell, D. G., Vedder, J. H. & McDougall, K.) 1–46 (1977).
5. Howell, D. G., McLean, H. & Vedder, J. H. *EOS* **61**, 949 (1980).
6. Vedder, J. H., Howell, D. G. & McLean, H. *Mem. Am. Ass. petrol. Geol.* (in the press).
7. Howell, D. G. & Vedder, J. H. *Soc. Economic Paleontologists and Mineralogists, (Pacific Sec.), Pacific Coast Paleogeography Symp.* Vol. 2 (eds Howell, D. G. & McDougall, K. A.) 107–116 (1978).
8. Champion, D. E., Gromme, C. S. & Howell, D. G. *Soc. Economic Paleontologists and Mineralogists, (Pacific Sec.), Field Trip Guide, Upper Cretaceous and Paleocene Turbidites, Central California Coast,* 53–56 (1981).
9. Irving, E. *Can. J. Earth Sci.* **16**, 669–694 (1979).
10. Hopson, C. A., Mattinson, J. M. & Pessagno, E. A. in *The Geotectonic Evolution of California* (ed. Ernst, W. G.) (Prentice-Hall, Englewood Cliffs, New Jersey, 1981).
11. Cox, A. in *The Geologic Timescale* (ed. Harland, W. B.) (Cambridge University Press, in the press).
12. Dickinson, W. R. *Rev. geophys. Space Phys.* **8**, 813–860 (1970).
13. Dickinson, W. R. *Am. J. Sci.* **272**, 551–576 (1972).
14. Dickinson, W. R. & Rich, E. I. *Bull. geol. Soc. Am.* **83**, 3007–3024 (1972).
15. Nilsen, T. H. & Clarke, S. H. *U.S. geol. Surv. Professional Pap.* **925**, 1–64 (1975).
16. Morgan, W. J. in *The Sea* (ed. Emiliani, C.) (Wiley, New York, 1980).
17. Page, B. M. *Am. J. Sci.* (in the press).

Implications of Palaeozoic phosphorites in the northern Sierra Nevada range

Robert J. Varga

Union Oil Research Center, PO Box 76, Brea, California 92621, USA

Marine phosphorites are known to occur within multiply deformed Ordovician/Silurian and Upper Devonian rocks of the northern Sierra Nevada range, California. Although the phosphatic sediments constitute a small percentage of total lithologies in the area, they provide unique information regarding the age and palaeogeographical setting of these rocks. The mode of occurrence of the Sierran phosphorites and their association with black shale and chert suggest that they may have formed along a continental margin which experienced strong equatorial or trade wind belt upwelling. Consideration of worldwide plate reconstructions for the early to middle Palaeozoic shows that the present western margin of North America was one of the few favourable sites for phosphate formation during this time period. Thus, the presence of marine phosphorites within Palaeozoic rocks of the northern Sierra Nevada is consistent with arguments which suggest that these rocks were deposited near the western margin of North America. Although considerable margin-parallel translations cannot be ruled out for these rocks, they do not appear to constitute a far-travelled microplate derived from a distant, trans-oceanic site.

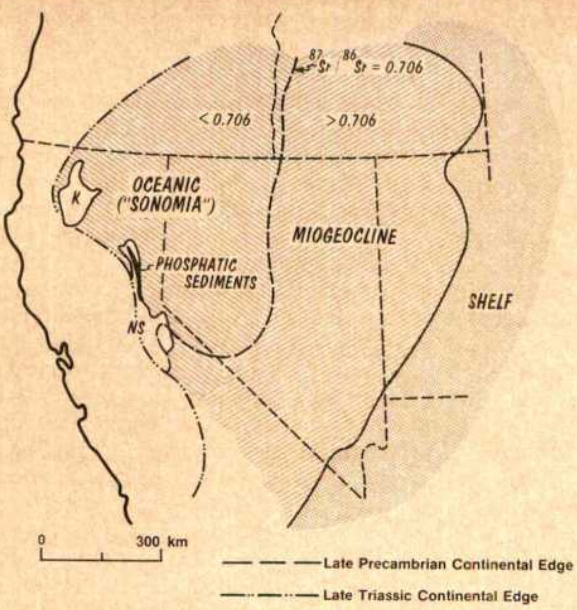


Fig. 1 Sketch map showing lower to middle Palaeozoic petrotectonic assemblages of the US Cordillera. Generalized outcrops of Palaeozoic strata in the northern Sierra Nevada (NS) and eastern Klamath Mountains (K) are shown with no pattern within 'oceanic' assemblage. Data from refs 19, 38, 44.

Phosphorite occurs within the pre-late Devonian (in part Ordovician/Silurian) Shoo Fly Complex (SFC) and late Devonian Elwell and Sierra Buttes Formations (EF and SBF, respectively) of the northern Sierra Nevada range. These formations and their phosphatic components crop out in an elongate, north-west-trending belt more than 150 km in length (Fig. 1). The SFC comprises predominantly siliciclastic sediments, in part of continental derivation[1], as well as minor limestone, dolomite and mélange[2,3]. Black shale, chert and intermediate to silicic volcanics of the SBF and EF[4] overlie the SFC with angular discordance[5–7] All three formations are multiply deformed[7–9] and were metamorphosed to lower greenschist facies at ~150 Myr (ref. 10). The SBF and EF constitute components of a middle Palaeozoic to early Triassic volcanic arc terrane which extends to the eastern Klamath Mountains of northern California[11].

Metamorphism and intense deformation has hindered interpretation of the depositional environments of Palaeozoic strata in the northern Sierra Nevada. As discussed below, such information is critical to the evaluation of their pre-Mesozoic palaeogeographical position. Thus, the presence of a unique sedimentary feature, such as phosphorite, assumes heightened significance. It is, therefore, of interest to compare the Sierran phosphorites with well-documented occurrences of marine phosphates about which some palaeoenvironmental generalizations can be made.

Phosphate within Palaeozoic rocks of the northern Sierra Nevada occurs principally as nodules, and to a lesser degree as lenses and angular fragments within foliated black shale and chert. Phosphate nodules (Fig. 2*a*) are generally round to sub-round and vary in size from ~0.5 to 2.5 cm. The nodules are black on fresh surface and weather to a light grey to white colour. The porous nodule surfaces often show a crude concentric pattern with inner layers composed of darker, more porous material than outer layers. Phosphatic lenses commonly accompany phosphate nodules (Fig. 2*b*). These lenses are up to 2.5 cm thick and 1 m in length. Angular fragments of phosphate, up to 0.5 cm in length, are ubiquitous associates of both phosphate nodules and lenses.

Phosphate nodules are composed of dark brown, cryptocrystalline, isotropic collophane (Fig. 2*c*) with inclusions of silica and pyrite. Slight differences in colour, transparency and quartz content define a crude, discontinuous banding towards the outer boundaries of most nodules. Every nodule examined contains a characteristic black, iron-rich outer surface which is up to 0.6 mm wide (Fig. 2*c*). Silica occurs as fine-grained intergrowths within the collophane matrix, as irregular masses up to 0.2 mm in size, and as recognizable biogenic remains (Fig. 2*d*). Well-preserved sponge spicules and radiolaria have been recovered by HCl dissolution of nodules from the EF[12] and SFC[7].

The average of four microprobe analyses of nodule interiors from the SFC is as follows: $P_2O_5 = 39.96\%$, $F = 3.56\%$, $CaO = 52.04\%$, $SiO_2 = 1.71\%$, $Cl{=}Sr{=}Na_2O{=}SO_3{=}MgO{=}K_2O <$ 500 p.p.m., total = 97.27% (the slightly low total could not be accounted for and presumably represents a small amount of structural OH^-, instrument calibration error, or both). CaO/P_2O_5 and F/P_2O_5 ratios in the SFC nodules have values of 1.30 and 0.09, respectively, which are close to the ratios

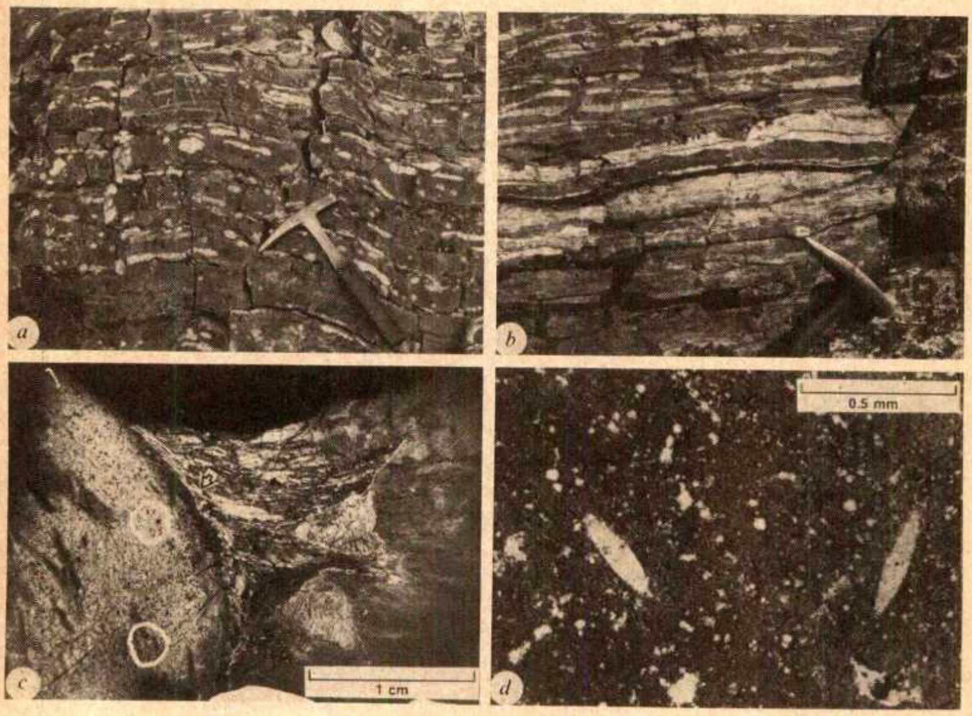


Fig. 2 *a*, Phosphate nodules (white areas) within black chert of the Elwell Formation. *b*, Phosphate lenses (white areas) and nodules (near hammer tip) within Shoo Fly Complex. *c*, Negative photograph of thin section showing the boundary between an undeformed phosphate nodule from the SFC (left side of photograph) and highly deformed black shale matrix. Note faint concentric banding within nodule and dark, iron-rich coating (light band in photo) on outer nodule surface (tip of arrow). Circled areas are probe sites. *d*, Sponge spicules(?) within SFC nodule.

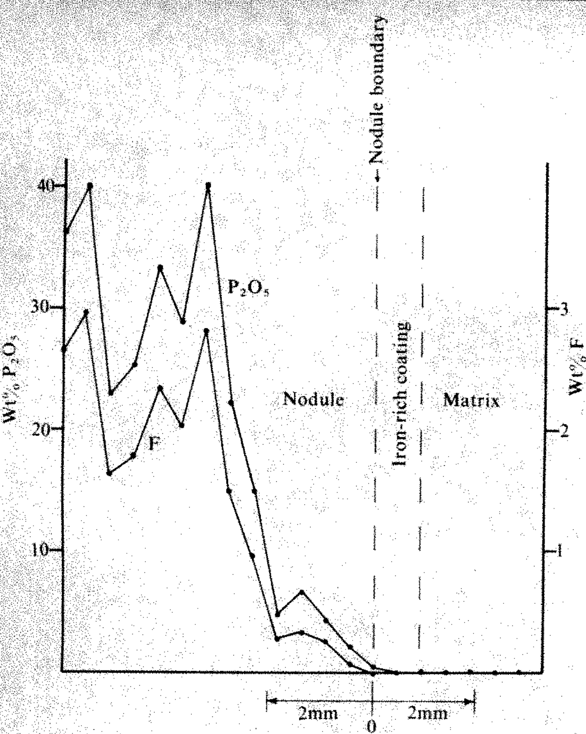


Fig. 3 Microprobe data from a 9.5-mm traverse across an SFC nodule–matrix boundary. Peaks and valleys in P_2O_5 and F contents show the relative abundances of apatite within the nodule. Areas of high apatite concentration correspond to layers of low included silica content. Iron-rich coating is seen as the dark band in Fig. 2*c* which separates matrix from nodule.

expected (1.32, 0.09) in unsubstituted fluorapatite, $Ca_{10}(PO_4)_6F_2$. CO_2 content was estimated using Gulbrandsen's[13] X-ray peak-pair method. By this method, CO_2 content of two SFC nodules and one EF nodule average $0.38\pm0.56\%$ and $0.74\pm0.56\%$, respectively. These estimates are consistent with a nearly ideal fluorapatite composition of the Sierran nodules.

The composition of the Sierran nodules is unusual in that most marine apatites contain 3–6% CO_3^{-2} substituted into the PO_4^{-3} site with the charge balanced by a variety of ions[14]. Post-depositional weathering, as a means of removing CO_2 and other components from the apatite structure[15], seems an unlikely mechanism to explain the apatite chemistry of the Sierran nodules. Outcrops from which the samples were collected for this study were quite fresh, having been scoured by alpine glaciers during the Pleistocene. Thus, the present chemistry of the nodules may represent either their original chemistry or the effects of chemical exchange during Jurassic dynamothermal metamorphism.

The origin of the concentric banding in Sierran phosphate nodules (Fig. 2*c*) is of interest because of the observation that many recent phosphorites contain films and layers of manganese[16]. Banding within an SFC nodule was investigated optically and by a 9.5-mm microprobe scan (Fig. 3) for P, F and Mn from the interior of the nodule to its outer edge. Weight percentages of both P_2O_5 and F, and thus apatite, show peak concentrations 3.3 and 5.6 mm from the outer boundary of the nodule with a drop in concentration in between. From 3.3 mm to the nodule boundary the concentration of apatite steadily decreases to zero with effectively no phosphorous or fluorine in the matrix. Mn was not present above detection limits (500 p.p.m.), anywhere along the microprobe traverse. X-ray fluorescence of the dark outer coating surrounding the nodule (Fig. 2*c*) showed it to be iron-rich with small amounts of titanium, calcium and potassium. Banding within nodule interiors thus corresponds to differences in apatite concentration. It is apparent optically that an increase in intergrown silica generally corresponds to decreases in apatite content. The sharp decrease in apatite content towards the nodule boundary is interpreted as indicating a steady drop in accretion of apatite during the waning stages of nodule growth.

The pre-Mesozoic position of Palaeozoic rocks of the northern Sierra Nevada with respect to North America has caused some debate[17]. The question is whether these rocks represent autochthonous or parautochthonous terranes which have been associated with the western margin of North America since the early Palaeozoic[7,11,18] or whether they represent components of a far-travelled island arc terrane, Sonomia (Fig. 1), which was sutured to North America in the Triassic[19]. Lack of faunal control and the absence of relevant palaeomagnetic data[10] preclude conclusive statements regarding the site of origin of these rocks. Thus, less direct petrological, sedimentological and structural arguments must be constructed[1,7].

The presence of marine phosphorites within the rock record of a particular region may have palaeogeographical implications if some analogy can be drawn between ancient and recent deposits. It is evident that many marine environments favourable to phosphate deposition in ancient oceans do not exist at present[13,16]. However, while the oceanic and biogeochemical conditions which existed during deposition of many ancient phosphorites remain largely unexplained, several recurrent observations lead to the following conclusions:

(1) When corrected for palaeomagnetic position, most recent and ancient phosphorites are found to have formed along the margins of continents in areas of oceanic upwelling at latitudes lower than 40°–50° (refs 20, 21).

(2) Phosphorites which form along the west coasts of continents ('west-coast type' deposits) in areas of trade wind belt

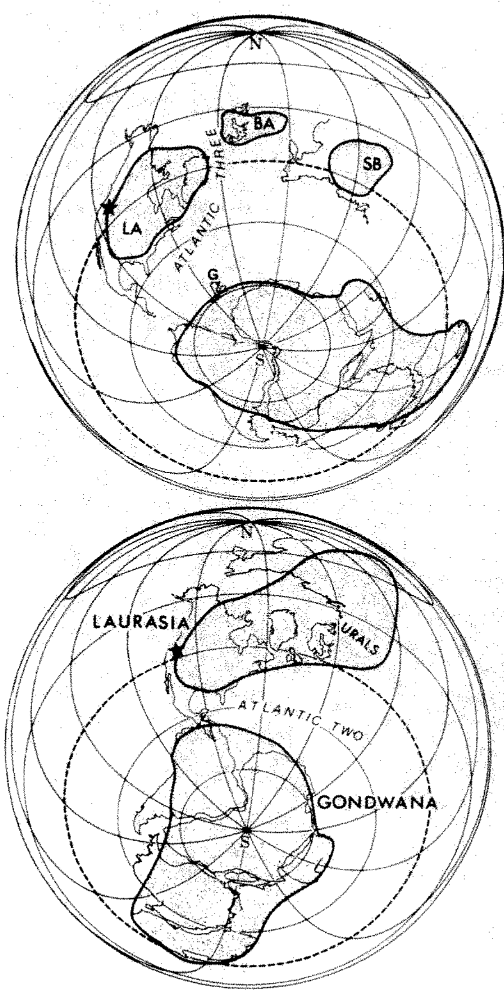


Fig. 4 Palaeomagnetic plate reconstructions for *a*, late Ordovician and *b*, late Devonian from ref. 34. ★, Present location of Sierran phosphorites with respect to North American Cordillera. Palaeoequators are shown as dashed lines.

upwelling[22-24] and those formed along continental margins bordering narrow, east–west seaways ('Tethyan-type' deposits)[25] in areas of equatorial upwelling[26] are typically associated with black shale and chert. Phosphate within west coast- or Tethyan-type deposits is generally nodular to pelletal in form.
(3) 'Florida-type'[27] or 'Chatham Rise-type'[28] phosphate deposits, which form along the east-facing margins of continents, are not generally associated with chert and black shale but typically occur within sandstone, limestone and dolomite[27,29,30]. Phosphate within Florida-type deposits is generally found in the form of pellets, pavements, sands and cobbles which show abundant evidence of enrichment by weathering, redeposition and replacement of limestone[30-32].

In form and lithic association phosphate deposits within Palaeozoic strata of the northern Sierra Nevada closely resemble west coast- or Tethyan-type deposits. Further, continentally derived sandstones within the SFC[1] imply relative proximity of the Sierran block to a continental mass during the early Palaeozoic. It is thus postulated that the Sierran phosphorites were deposited along the west-facing coast of a relatively large continental mass or, possibly, in a narrow east–west seaway between two continents during the early to middle Palaeozoic. By analogy with the world-wide distribution of modern and ancient phosphorites[20,21], the site of deposition would have been located at a latitude lower than 40°–50°. Water depths may have been in the range of 100–500 m (ref. 33).

Figure 4*a*, *b* shows the world-wide plate configuration as determined by Morel and Irving[34] for the general time period of phosphate deposition in the northern Sierra Nevada, approximately late Ordovician to late Devonian. The maps show that the west coast of North America was one of the few continental margins situated favourably throughout the early to middle Palaeozoic for formation of west coast-type phosphorites as most of the Gondwana continents were clustered about the South Pole. Other continental margins in positions favourable for west coast-type phosphate formation during the entire time period were the present north-west coast of Australia and northern India. The Palaeozoic stratigraphy and structural development of north-west Australia[35] and northern India[36] bear little resemblance to that of the northern Sierra Nevada[4,7,8]. Thus, neither of these sites is believed to represent a reasonable source terrane for a Sierran microplate. The maps also show that no narrow, equatorial seaways existed during the early to middle Palaeozoic, an observation consistent with the plate reconstructions of Scotese and others[37].

The present position of Sierran phosphorites near the western edge of the Cordillera is consistent with the palaeogeography of the North American continent during the early to middle Palaeozoic. If the northern Sierra Nevada was in its approximate position relative to North America during the early to middle Palaeozoic then the phosphorites would have lain at approximately lat. 5°–10° N during this time period. Schweickert and Snyder[38] concluded that an ocean as wide as the modern Atlantic may have existed off western North America by the Middle Ordovician assuming spreading rates comparable with those along the Mid-Atlantic Ridge and an upper Proterozoic (<850 Myr) rifting event[39]. The west-facing Palaeozoic margin of North America, at low latitudes and opposite an ocean sufficiently wide to develop circulatory gyres, was an ideal location to develop phosphatic sediments. It is thus not surprising to find phosphorites within Sierran strata if indeed they are indigenous to North America.

The hypothesis that Palaeozoic sediments of the northern Sierra Nevada were deposited on or near the Cordilleran margin of North America during the early to middle Palaeozoic has important tectonic implications. Specifically, this hypothesis is inconsistent with models which suggest that the northern Sierra Nevada represents part of a wholly exotic island arc terrane which was unrelated to North America until accretion during the Permo–Triassic Sonoma Orogeny[19]. Considerable translation of the northern Sierra Nevada along the continental margin[19] cannot be ruled out, however, as the entire west coast of North America could have received phosphate deposition (Fig. 4*a*, *b*). Though not conclusive, the present position of the northern Sierra Nevada within an area favourable for phosphate deposition during the early to middle Palaeozoic suggests a less complicated model wherein the Sierran phosphorites were deposited along the western margin of North America and have since experienced only limited translations. This model is consistent with the observations of other authors who have noted broad tectonic and stratigraphical similarities between Palaeozoic rocks of the Great Basin and the Sierra Nevada[1,4,7,11,38,40-43].

I thank G. Bond for discussions, M. Bell, P. Fenn and E. Goldish for analytical work, Union Oil Company for assistance in manuscript preparation, and W. Burnett, G. A. Crawford and E. Irving for review. Work was supported by NSF grant EAR 76-14270 awarded to E. Moores and G. Bond, University of California at Davis.

Received 2 December 1981; accepted 19 March 1982.

1. Bond, G. C. & Devay, J. *J. Geol.* **88**, 285–308 (1980).
2. Clark, L. D. *Prof. Pap. U.S. geol. Surv.* **923**, 26 (1976).
3. Schweickert, R. A. *Abstr. Prog. geol. Soc. Am.* **6**, 251 (1974).
4. D'Allura, J. A., Moores, E. M. & Robinson, L. *Pacif. Coast Paleogeogr. Symp. Soc. Econ. Paleontol. Miner.* **1**, 421–440 (1977).
5. Durrell, C. & D'Allura, J. A. *Bull. geol. Soc. Am* **88**, 844–852 (1977).
6. Schweickert, R. A. & Girty, G. H. *Abstr. Prog. geol. Soc. Am.* **13**, 105 (1981).
7. Varga, R. J. & Moores, E. M. *Geology* **9**, 512–518 (1981).
8. Girty, G. H. & Schweickert, R. A. *Abstr. Prog. geol. Soc. Am.* **11**, 79 (1977).
9. Varga, R. J. *Abstr. Progr. geol. Soc. Am.* **11**, 133 (1977).
10. Hannah, J. L. & Verosub, K. L. *Geology* **8**, 520–524 (1980).
11. Burchfiel, B. C. & Davis, G. A. *Am. J. Sci.* **275A**, 363–396 (1975).
12. Devay, J. & Stanley, E. *Abstr. Prog. geol. Soc. Am.* **11**, 412 (1979).
13. Gulbrandsen, R. A. *Prof. Pap. U.S. geol. Surv.* **700-B**, 9–13 (1970).
14. Bentor, Y. K. *Spec. Publ. Soc. Econ. Paleont. Miner.* **29**, 3–18 (1980).
15. McArthur, J. M. *Spec. Publ. Soc. Econ. Paleont. Miner.* **29**, 53–60 (1980).
16. Bentor, Y. K. in *Rep. Marine Phosphatic Sed. Workshop*, 29 (1979).
17. Dickinson, W. R. *Pacif. Coast. Paleogeogr. Symp. Soc. Econ. Paleont. Miner.* **1**, 137–155 (1977).
18. Churkin, M. Jr. *Spec. Pub. Soc. Econ. Paleont. Miner.* **19**, 174–192 (1974).
19. Speed, R. C. *Geology* **87**, 279–292 (1979).
20. Sheldon, R. P. *Prof. Pap. U.S. GEOL. Surv.* **501C**, 106–113 (1964).
21. Cook, P. J. & McElhinny, M. W. *Econ. Geol.* **74**, 315–330 (1979).
22. McKelvey, V. E., Swanson, R. W. & Sheldon, R. P. *19th int. Geol. Congr.* **11**, 45–65 (1953).
23. Grower, H. D. & Madsen, B. M. *Prof. Pap. U.S. geol. Surv.* **501D**, 79–85 (1964).
24. Cheney, T. M., McClellan, G. H. & Montgomery, E. S. *Econ. Geol.* **74**, 232–259 (1979).
25. Sheldon, R. P. *Econ. Geol.* **59**, 1159–1175 (1964).
26. Sheldon, R. P. *Spec. Pap. Soc. Econ. Paleont. Miner.* **29**, 239–247 (1980).
27. Cathart, J. B. *U.N. Res. Dev. Ser.*, **32**, 403–405 (1969).
28. D'Anglejan, B. F. in *Rep. Marine Phosphatic Sed. Workshop*, 17 (1979).
29. McKelvey, V. E. *Bull. U.S. geol. Surv.* **1252D**, 1–22 (1967).
30. Riggs, S. R. *Econ. Geol.* **74**, 195–220 (1979).
31. Manheim, F. T., Pratt, R. M. & McFarlin, P. F., *Spec. Publ. Soc. Econ. Paleont. Miner.* **29**, 117–137 (1980).
32. Cullen, D. J. *Spec. Publ. Econ. Paleont. Miner.* **29**, 139–148 (1980).
33. Burnett, W. C. in *Proc. Fertilizer Raw Materials Workshop*, 120–144 (1979).
34. Morel, P. & Irving, E. *J. Geol.* **86**, 535–561 (1978).
35. Veevers, J. J. *J. geol. Soc. Aust.* **14**, 253–271 (1967).
36. Gansser, A. *Geology of the Himalayas*, 289 (Interscience, New York, 1964).
37. Scotese, C., Bambach, R. K., Barton, C., Van der Voo, R. & Ziegler, A. *J. Geology* **87**, 217–277 (1979).
38. Schweickert, R. A., & Snyder, W. S. in *The Geotectonic Development of California*, 182–202 (Prentice-Hall, New Jersey, 1981).
39. Stewart, J. H. *Bull. geol. Soc. Am.* **83**, 1345–1360 (1972).
40. Moores, E. M. *Nature* **228**, 837–842 (1970).
41. Schweickert, R. A. *Nature* **260**, 586–591 (1976).
42. Schweickert, R. A. *Abstr. Prog. geol. Soc. Am.* **9**, 497 (1977).
43. Moore, J. N. & Foster, C. T. *Bull. geol. Soc. Am.* **91**, 37–43 (1980).
44. Burchfiel, B. C. & Davis, G. A. in *The Geotectonic Development of California*, 50–70 (Prentice-Hall, New Jersey, 1981).

Temperature control of oxygen-isotope fractionation of cultured planktonic foraminifera

Jonathan Erez

The Hebrew University of Jerusalem, The H. Steinitz Marine Biology Laboratory, Eilat, PO Box 469, Israel

Boaz Luz

The Hebrew University of Jerusalem, Department of Geology, Jerusalem, Israel

The palaeotemperature equation of Epstein *et al.*[1] relates the oxygen-isotope composition of mollusc carbonate shells to the temperature in which it has been deposited. This equation has been applied extensively to derive palaeotemperatures and

other palaeoenvironmental parameters for the Pleistocene epoch, based on the isotopic composition of planktonic foraminifera in deep sea cores[2–11]. Recently, doubt has been expressed as to whether planktonic foraminifera fractionate oxygen isotopes according to the known palaeotemperature equation[12–18], and thus the validity of many palaeoenvironmental studies has been questioned. We present here experimental data demonstrating that planktonic foraminifera of the species *G. sacculifer* deposit calcite skeletons that have an oxygen-isotope composition very close to that predicted by the well-known palaeotemperature equation[1].

The oxygen-isotope composition of planktonic foraminifera in deep sea cores is a major tool in Pleistocene research. It serves to determine palaeotemperatures, palaeosalinities, sea-level changes, continental climatic conditions and provides accurate stratigraphy for the Pleistocene epoch[2–11]. These studies are based on the assumption that oxygen isotopes in planktonic foraminifera shells are fractionated according to the equation of Epstein *et al.*[1] that was determined experimentally for molluscs. Emiliani[2] tested this assumption by comparing the 'isotopic temperature' (the temperature calculated based on the isotopic composition of calcium carbonate), of planktonic foraminifera from superficial sediments, to the actual temperatures of the overlying seawater. He found a good fit between these temperatures and concluded that planktonic foraminifera can be used for palaeotemperature studies. This approach has been repeated on material from sediments with similar conclusions[6,10,19,20]. However, isotopic temperatures calculated for living planktonic foraminifera that were collected by plankton net tows were almost always significantly higher than those observed in the water in which they were collected[12–18]. It has been proposed that these nonequilibrium isotopic compositions are caused by introduction of metabolic CO_2 into the carbonate skeletons, a phenomenon that has been reported in other carbonate depositing organisms, including benthic foraminifera[21,22] and in part was attributed to metabolic activity of symbiotic algae[21]. Attempts to resolve the problems using foraminifera collected in sediment traps did not reach unequivocal conclusions[18,23], and thus palaeoclimatic determinations based on oxygen-isotope compositions of planktonic foraminifera became uncertain.

The ability to culture planktonic foraminifera in the laboratory[24] opened the possibility of resolving this problem experimentally. A detailed account of the entire experiment will be reported elsewhere[25]. Briefly, young (roughly 10-day old) individuals of the planktonic foraminifer *G. sacculifer* were collected from surface water in the Gulf of Eilat using plankton nets. Each individual was cultured in an 100-ml Erlenmeyer flask with 80 ml of filtered seawater and was fed daily by one newly hatched brine shrimp. The Erlenmeyer flasks were kept in temperature controlled baths where the deviations from the desired temperature did not exceed 0.1 °C (based on three daily accurate thermometer readings, and a continuous monitoring of temperature on strip chart recorder). Salinities were monitored every 2 days to ensure that there was no water evaporation. The growth of the foraminifera was monitored daily using an inverted microscope. All individuals considered in this experiment reached maturity and went through reproduction by producing gametes. The average initial weight of these foraminifers was ~5 μg per individual and the final weight was ~45 μg per individual. Thus ~90% of the carbonate shells was deposited in controlled conditions. A correction for the original (10% weight) isotopic composition was made using control groups for which the isotopic composition was determined separately. Each experiment lasted roughly 2 weeks; the range of experimental temperatures was between 14 and 30 °C, and in each run there were 2 or 3 different temperatures. Because only individuals that went through gametogenic reproduction were used for the final isotopic analysis, there was no need to clean the protoplasm from the shell. During the process of gametogenesis[24], the entire organic matter has been extruded naturally from the shell and the skeleton that is left behind is white, organic matter free and very similar to material from sediments. Yet before analysis the individuals were crushed in ethanol, dried, vacuum roasted for 30 min at 450 °C. The analysis was carried out according to the method of Shackleton[26] using a VG602 Micromass spectrometer. The water isotopic composition was determined according to the procedure of Epstein and Mayeda[27].

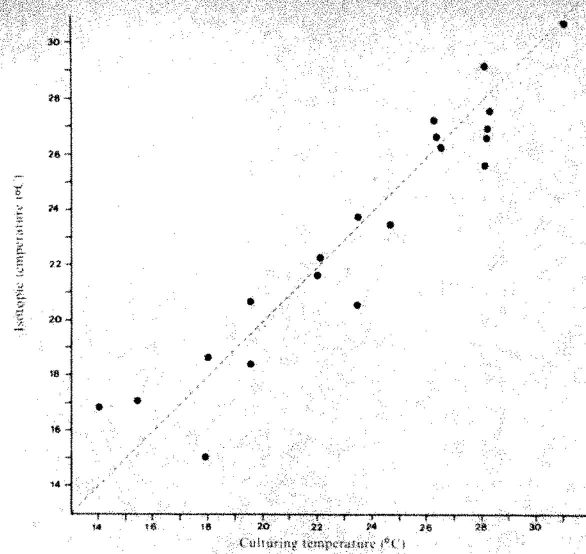


Fig. 1 The isotopic temperature calculated from the oxygen-isotope composition of the shell carbonate of the foraminifer *G. sacculifer* according to the modified equation of Epstein *et al.*[1] plotted against the actual temperature in which the foraminifers were cultured. The dashed line represents the ideal fit between the two temperatures having a slope of 1.0 and an intercept of 0. The best linear fit to the data is a line having slope and intercept of 1.01 and −0.07, respectively, having a correlation coefficient of $r = 0.95$. These experimental data demonstrate that *G. sacculifer* deposits its skeleton close to oxygen isotopic equilibrium, and that the common palaeotemperature equation for molluscs is applicable for this planktonic foraminifer.

The isotopic temperatures were calculated according to the equation of Epstein *et al.*[1] as modified by Craig[28]

$$pt = 16.9 - 4.2(\delta^{18}O_c - \delta^{18}O_w) + 0.13(\delta^{18}O_c - \delta^{18}O_w)^2$$

where pt is the isotopic temperature and $\delta^{18}O_c$ and $\delta^{18}O_w$ are the isotopic composition of the carbonate and seawater respectively.

The isotopic temperature and the actual culture temperature are compared in Fig. 1. A remarkable fit exists between the isotopic and the actual temperatures. The line of best fit using linear regression has the following parameters:

$$t = -0.07 + 1.01pt$$

where t and pt are the actual and the isotopic temperatures, respectively. The correlation coefficient (r) for this linear regression equals 0.95. Ideally, the intercept and the slope should have been 0.0 and 1.0, respectively. The values that we obtained (−0.07 and 1.01) are almost identical to those expected, and would cause only a slight deviation of <1% in palaeotemperature determinations in the range of 14–30 °C.

We therefore conclude that the planktonic foraminifer *G. sacculifer* fractionate oxygen isotopes with changing temperatures according to the modified equation of Epstein *et al.*[1] that was originally developed for molluscs. Accurate palaeotemperature determinations that are based on *G. sacculifer* isotopic composition should still take into consideration the isotopic composition of seawater and the vertical migration of this species during its ontogeny[12,23]. The nonequilibrium values that have been observed in plankton samples deserve further investigation, but it is not unlikely that they were caused by vacuum roasting of the samples before analysis, which was intended to remove organic contaminants.

Our experiment provides more data points to the palaeotemperature equation at the temperature range where it was lacking (14–30 °C) and substantiates the results of earlier palaeoenvironmental studies based on the oxygen-isotope composition of planktonic foraminifera.

We thank A. Szin and A. Shemesh for technical assistance.

Received 14 December 1981; accepted 15 March 1982.

1. Epstein, S., Buchsbaum, R., Lowenstam, H. A. & Urey, H. C. *Bull. geol. Soc. Am.* **64,** 1315–1326 (1953).
2. Emiliani, C. *Am. J. Sci.* **252,** 149–158 (1954); *J. Geol.* **63,** 538–578 (1955).
3. Emiliani, C. *J. Geol.* **74,** 102–126 (1966).
4. Shackleton, N. J. & Opdyke, N. D. *Quat. Res.* **3,** 39–55 (1973); *Geol. Soc. Am. Mem.* **145,** 449–464 (1976).
5. Imbrie, J., Van Donk, J. & Kipp, N. G. *Quat. Res.* **3,** 10–38 (1973).
6. Shackleton, N. J. & Vincent, E. *Mar. Micropaleont.* **3,** (1978).
7. Deuser, W. G., Ross, E. H. & Waterman, L. S. *Science* **191,** 1168–1170 (1976).
8. Vergnaud-Grazzini, C., Ryan, W. B. F. & Cita, M. B. *Mar. Micropaleont.* **2,** 353–370 (1977).
9. Williams, D. F., Thunell, R. C. & Kennett, J. P. *Science* **201,** 252–254 (1978).
10. Berger, W. H., Killingley, J. S. & Vincent, E. *Oceanol. Acta* **1,** 203–216 (1978).
11. Reiss, Z. *et al. Quat. Res.* **14,** 294–308 (1980).
12. Van Donk, J. thesis, Columbia Univ., New York (1970).
13. Van Donk, J. *Oceanic Micropaleontology* (ed. Ramsay, A. T. S.), 1345–1370 (Academic, New York, 1977).
14. Shackleton, N. J., Wiseman, J. D. & Buckley, H. A. *Nature* **242,** 177–179 (1973).
15. Kahn, M. I. *Oceanol. Acta* **2,** 195–208 (1979).
16. Kahn, M. I. & Williams, D. F. *Paleogeogr. Paleoclimatol. Paleoecol.* **33,** 47–69 (1981).
17. Duplessy, J. C., Be, A. W. H. & Blank, P. L. *Paleogeogr. Paleoclimatol. Paleoecol.* **33,** 9–46 (1981).
18. Deuser, W. G., Ross, E. H., Hemleben, C. & Spindler, M. *Paleogeogr. Paleoclimatol. Paleoecol.* **33,** 103–127 (1981).
19. Lidz, B., Kehm, A. & Miller, H. *Nature* **217,** 245–247 (1968).
20. Savin, S. M. & Douglas, R. G. *Bull. geol. Soc. Am.* **84,** 2327–2342 (1973).
21. Erez, J. *Nature* **273,** 199–202 (1978).
22. Vinot-Bertuille, A. C. & Duplessy, J. C. *Earth planet Sci. Lett.* **18,** 247–252 (1973).
23. Erez, J. & Honjo, S. *Paleogeogr. Paleoclimatol. Paleoecol.* **33,** 129–156 (1981).
24. Be, A. W. H. *et al. Micropaleontology* **23,** 155–179 (1977).
25. Erez, J. & Luz, B., *Geochim. cosmochim. Acta* (submitted).
26. Shackelton, N. J. *Colloqu. int. cent. natn. Rech. scient.* No. 219, 203–209 (1973).
27. Epstein, S. & Mayeda, T. *Geochim. cosmochim. Acta* **4,** 213–224 (1953).
28. Craig, H. *Proc. Spoleto Conf. on Stable Isotopes in Oceanographic Studies and Paleotemperatures* Vol. 3 (ed. Tongiorgi, E.) (1965).

IQ in Japan and the United States shows a growing disparity

Richard Lynn

Department of Psychology, The New University of Ulster, Coleraine, Co. Londonderry BT52 1SA, UK

Evidence from 27 samples indicates that the mean IQ in Japan is higher than in the United States by around one-third to two-thirds of a standard deviation. Analysis of results from the standardization in Japan in 1975 of the new revised version of the American Wechsler Intelligence Scale for Children shows that the Japanese–American disparity in mean IQ has increased during the twentieth century. Among the younger generation the mean Japanese IQ is approximately 111.

Several studies have indicated that the mean IQ is higher in Japan than in the United States and other advanced western nations[1,2]. The Japanese–American disparity has varied in different samples from 2 to 15 IQ points. The existing evidence suggests that the disparity is generally smallest among samples born in the earlier decades of the century, and greatest among young samples born more recently. Thus the mean IQ in Japan may have risen relative to that in the United States, producing an increasing disparity between the two populations.

Evidence to determine whether this is the case can be derived from the recent standardization in Japan of the revised version of the Wechsler Intelligence Scale for Children[3]. This test was constructed and standardized in the United States in the early 1970s and was standardized in Japan in 1975. Five of the performance subtests were retained unaltered in the Japanese standardization—Picture Arrangement, Object Assembly, Coding, Mazes and Block Design (in Block Design the last item in the Japanese version has been made slightly more difficult, thus making the test harder for the Japanese). The Digit Span subtest was also unchanged in the Japanese standardization, as in the previous standardizations of the Wechsler tests.

In Japan, the test was standardized on a representative sample of the child population consisting of 100 subjects at each age level from 6–16 yr. The standardization procedure in Japan followed that in the United States in matching the sample for socio-economic status to the total population. The mean raw scores of the Japanese samples at each age can be read off from the Japanese manual, and then converted to American scaled scores and the corresponding IQs ascertained from the American manual. Table 1 gives the mean Japanese scaled scores on the five performance subtests and on Digit Span, and the mean performance IQ derived from the five subtests. The overall mean performance IQ of this Japanese sample is 111 (relative to a mean American IQ of 100). 1,100 Japanese and 2,200 Americans were tested and the difference in mean IQ is clearly statistically significant. For individual tests for each year, a difference of one scaled point is statistically significant at the 0.01 probability level. The result evidently confirms previous estimations indicating that the Japanese mean IQ is higher than that in the United States.

Examination of the scores obtained by the Japanese on the six subtests given in Table 1 shows that the Japanese superiority is most pronounced in the tests of block design, mazes, picture arrangement and object assembly. The Japanese are less good on the digit span and coding tests. These results are broadly similar to those found in the Japanese standardization of the WISC and the WAIS. The consistency of this pattern of results suggests that the Japanese may be particularly strong on spatial ability.

Combining the results from the WISC-R with those of previous studies we have IQ data for 27 Japanese cohorts spanning almost seven decades. The series starts with the results for the WAIS for which the Japanese standardization sample was born around 1910 and concludes with the 6-yr-olds in the WISC-R born in 1969. Table 2 gives the mean IQs of the 27 cohorts in relation to their year of birth. It is clear that the Japanese cohorts born earlier in the century (1910–45) have a mean IQ of 102–105, whereas those born from 1946 to 1969 have a mean of 108–115. This suggests that the mean Japanese IQ has been rising relative to the American during the twentieth century. The secular trend is shown in Fig. 1, in which mean

Table 1 Japanese mean WISC-R scaled scores and IQs based on US norms

Age	Digit span	Picture arrangement	Block design	Object assembly	Coding	Mazes	Performance IQ
6	10.5	11.8	13.9	11.6	9.0	13.2	112
7	10.5	11.7	14.2	11.7	7.7	13.0	111
8	10.7	10.8	14.4	11.5	9.4	12.7	112
9	9.8	12.0	14.9	11.0	9.2	12.0	112
10	10.7	11.3	14.7	11.7	9.2	11.7	111
11	10.7	11.0	14.9	11.7	9.3	11.5	111
12	10.7	11.0	14.3	11.7	9.7	11.8	112
13	10.0	11.0	14.1	11.7	9.9	11.3	111
14	10.8	11.0	14.2	11.0	9.8	12.0	111
15	10.5	11.0	14.2	11.0	9.8	11.0	109
16	11.5	11.0	13.5	11.0	9.3	10.0	106

0028–0836/82/200222—02$01.00

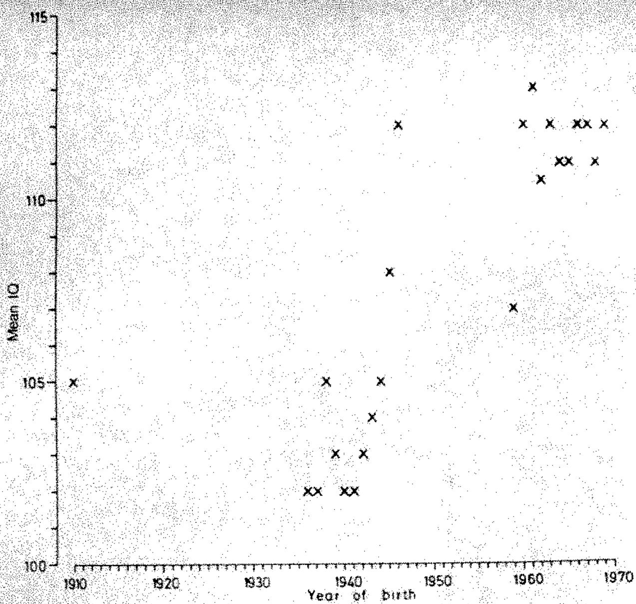


Fig. 1 Mean IQs of 23 Japanese cohorts showing a secular increase relative to a mean IQ of 100 in the US.

IQs for the same cohort are averaged to give a single mean (this applies to the four cohorts born during 1959–62).

The significance of the rising trend of the mean IQ in Japan can be tested from the correlation between mean IQ and year of birth. The product moment correlation is +0.76 and is statistically significant at $P<0.01$. The greatest discrepancy is in the value for 1910, which was derived from the WAIS and is based on only two of the performance subtests and digit span. It may therefore be subject to error. If this value is discounted, the cohorts born in the years 1936–46 have a mean IQ of 104.34, while those born in the years 1959–46 have a mean IQ of 111.20. The value of χ^2 for this difference is 10.83 and is statistically significant at $P<0.01$ (Yates' correction is applied).

Thus over the course of a generation the mean IQ in Japan has risen by ~7 IQ points. It seems doubtful whether a rise of this magnitude could be accounted for by a change in the genetic structure of the population. Instead, the explanation probably lies largely in environmental improvements. Comparison of the WISC and the WISC-R shows that the increase in IQ was present among 6-yr-olds. This indicates that it cannot be explained in terms of improvements in education and must be attributed to effects taking place before the age of six. Improvements in health and nutrition may be involved as it has been shown that the birth weight of Japanese babies has increased over the middle decades of the century[4].

Other advanced western nations, such as Britain, France, Belgium, Germany, Australia and New Zealand, all have mean IQs approximately the same as that in the United States[5]. At 111, the mean IQ in Japan is the highest recorded for a national population by a considerable order of magnitude. The effects of the high Japanese IQ of 111 can be illustrated as follows. Whereas Americans and Europeans have ~2% of their populations with IQs over 130, the Japanese have ~10% at this level. Among the population as a whole, 77% of Japanese have a higher IQ than the average American or European. Since intelligence is a determinant of economic success, as it is of success in many other fields, the Japanese IQ advantage may have been a significant factor in Japan's outstandingly high rate of economic growth in the post-World War II period.

Table 2 Mean IQs of 23 Japanese cohorts and their year of birth

Year of birth	Mean IQ	Test	Year of birth	Mean IQ	Test
1910	105	WAIS	1959	108	WPPSI
1936	102	WISC		106	WISC-R
1937	102	WISC	1960	115	WPPSI
1938	105	WISC		109	WISC-R
1939	103	WISC	1961	115	WPPSI
1940	102	WISC		111	WISC-R
1941	102	WISC	1962	110	Kyoto
1942	103	WISC		111	WISC-R
1943	104	WISC	1963	112	WISC-R
1944	105	WISC	1964	111	WISC-R
1945	108	WISC	1965	111	WISC-R
1946	112	WISC	1966	112	WISC-R
			1967	112	WISC-R
			1968	111	WISC-R
			1969	112	WISC-R

Received 30 November 1981; accepted 26 March 1982.

1. Lynn, R. *Bull. Br. psychol. Soc.* **30**, 69 (1977).
2. Lynn, R. & Dziobon, J. *Person. Indiv. Diff.* **1**, 95 (1980).
3. Kodama, H., Shinagowa, F. & Motegi, M. *Wechsler Intelligence Scale for Children—Revised* (Nihon Bunka Kagakusha, 1978).
4. Gruenwald, P., Funakawa, H., Mitani, S., Nishimura, T. & Takeuchi, S. *Lancet* **i**, 1026–1028 (1967).
5. Lynn, R. in *Human Variation* (eds Osborne, R. T., Noble, G. E. & Weyl, N.) (Academic, New York, 1978).

Light experience and asymmetry of brain function in chickens

Lesley J. Rogers

Pharmacology Department, Monash University, Wellington Road, Clayton, 3168 Victoria, Australia

Asymmetry of brain function has been demonstrated for several non-human species[1], most clearly in the avian brain[2,3]. In the chicken, a single treatment of the left forebrain hemisphere during the first week of life after hatching with either cycloheximide or the putative neurotransmitter, glutamate, subsequently causes retarded visual and auditory learning, and elevated attack and copulation responses[3–5]. Similar treatment of the right hemisphere is without effect. The cellular mechanisms by which these pharmacological agents may alter brain development and so reveal lateralization have already been discussed in detail[6,7]. It is commonly assumed that lateralization of brain function in humans and other species is inherited either genetically[1,8] or cytoplasmically[9]. However, Rogers and Anson have suggested previously[3] that light experience may have an important role in establishing lateralization in the chicken forebrain, because after day 17 of incubation the embryo is oriented in the egg such that the left eye is occluded by the chicken's wing and body while the right eye is next to the air sac and exposed to light input (refs 10, 11 and J. V. Zappia and L.J.R., in preparation). Because the optic nerves decussate completely and most of the information reaching each tectum is processed by its ipsilateral forebrain hemisphere, it is possible that light entering the right eye stimulates developmental processes in the left hemisphere in advance of the right[3]. If so, chickens hatched from eggs incubated in darkness should fail to show functional asymmetry of the forebrain. I now report that this is indeed the case for asymmetrical control of attack and copulation.

Because all previous experiments demonstrating asymmetrical control of attack and copulation have used only male chickens hatched from eggs exposed to light during incubation[4,5], it was first necessary to ascertain that the same phenomenon occurred in females also hatched from eggs exposed to light, since chicks hatched in my laboratory were not sexed. As found for males, glutamate (500 nmol in 5 μl) treatment of the left hemisphere of female chicks on day 2 after hatching caused elevated attack and copulation (see ref. 4 for details of method), whereas those treated similarly in the right hemisphere behaved the same as untreated controls (Fig. 1).

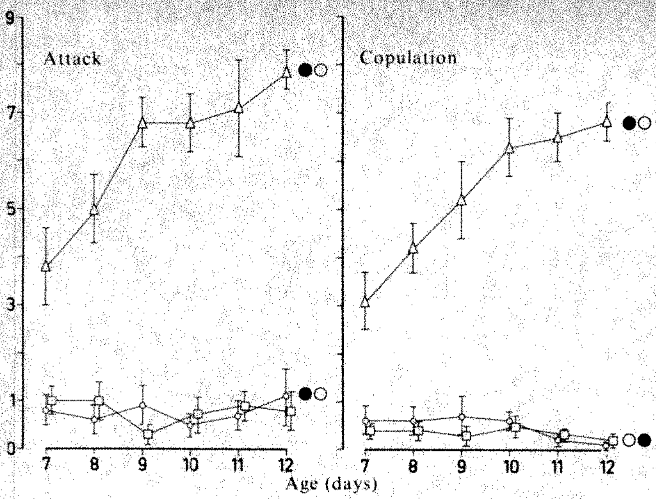


Fig. 1 Female chicks, supplied by a hatchery in which the eggs are exposed to light during incubation, have been tested for attack and copulation from days 7 to 12 of life. Daily mean scores with standard errors are presented. The behaviours are scored according to a rank ordering procedure (0–10) in standard tests using the moving hand as a stimulus[4]. Scores plotted on the y-axis rank from no response (0) to sparring with attack leaps for attack, and treading with pelvic thrusting for copulation. △ Represents females treated with glutamate in the left hemisphere, and on every day of testing their scores are significantly ($P<0.002$, U-tests) elevated above those for females treated in the right hemisphere (○), and untreated females (□). The pair of circles next to each graph represents the hemispheres, the hemisphere treated with glutamate being solid black. The other hemisphere is treated with normal saline. There were 8–10 animals in each group.

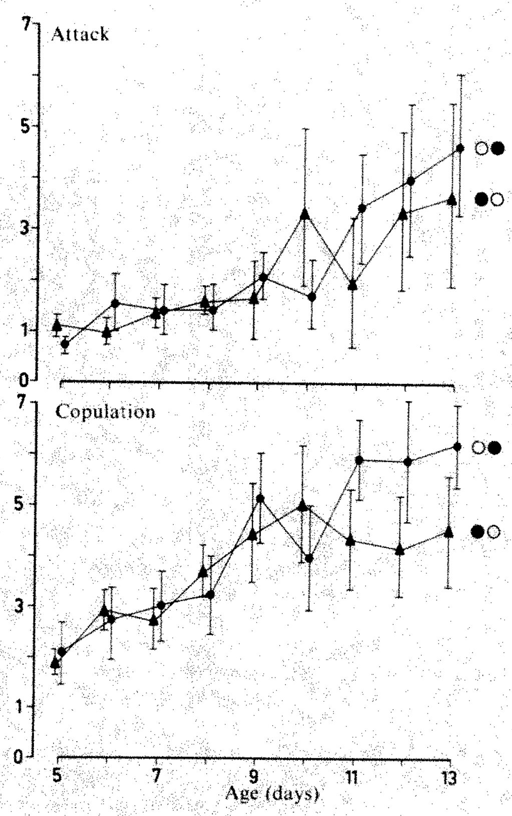


Fig. 2 Attack and copulation scores for chicks hatched from eggs incubated completely in darkness. Symbols as in Fig. 1; ▲, glutamate treatment in the left hemisphere ($n=15$); ●, glutamate treatment in the right ($n=10$).

Male and female chickens hatched from eggs incubated in darkness and treated with 500 nmol of glutamate on day 2 after hatching show no asymmetry of function for attack and copulation (Fig. 2). As the mean scores in Fig. 2 demonstrate, the chicks in both the left and right hemisphere treatment groups showed some elevation of attack and copulation scores.

In the chicken the first visually evoked tectal responses occur on day 17 of incubation. An electroretinogram of embryonic form can be detected and the first responses to light can be measured on day 18, and on day 19 the electroretinogram reaches its post-hatch form and light-evoked potentials are detected in the contralateral forebrain hemisphere[11,12]. Thus, it is during the last 5 days of incubation that light exposure should affect brain asymmetry, and this is in fact the case. In a matched control experiment, one batch of eggs was incubated in darkness for the first 16 days of incubation and then exposed to light of intensity varying between 100 and 500 lx during the last 5 days; another batch received the reverse treatment. The former treatment causes lateralization of atttack and copulation, the latter does not (see Fig. 3).

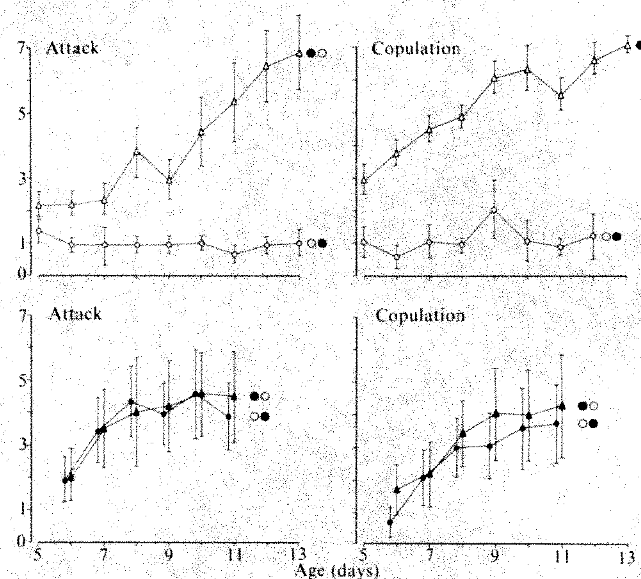


Fig. 3 The two graphs at the top present data for chicks hatched from eggs which were incubated in darkness for the first 16 days of incubation, and light from a heating bulb and overhead lamps (100–500 lx) during the last 5 days. The graphs at the bottom present data for eggs incubated in the reverse conditions ($n=8$–10 for each group). △, ▲, Treated with glutamate in the left hemisphere; ○, ●, treated in the right hemisphere.

In fact, 2 h of light exposure on day 19 of incubation are sufficient to establish asymmetrical control of attack and copulation (see Fig. 4). A sensitive period for this effect seems to have passed by the time the chick pips the egg shell on day 20 because no asymmetry for attack and copulation was present in chickens hatched from eggs exposed to light from the time of pipping onwards (~24 h).

As shown by the standard errors in the figures, the data for those groups of chickens hatched from eggs which had not received light exposure during the sensitive period were more variable than those of the lateralized groups ($P\leq 0.001$ for a comparison of the attack and the copulation scores on days 11 or 12 of all the lateralized versus non-lateralized chicks treated in the left hemisphere, and $P<0.05$ for both attack and copulation scores in a similar lateralized versus non-lateralized comparison of all those treated in the right hemisphere; Moses tests[13], $n=38$–48 for each group). Examination of the individual scores revealed that approximately half the chicks hatched from eggs kept in darkness during the sensitive period and then treated in the left hemisphere had consistently high scores for attack and copulation clustered around 7.0, while the other half had consistently low scores clustered around 1.0,

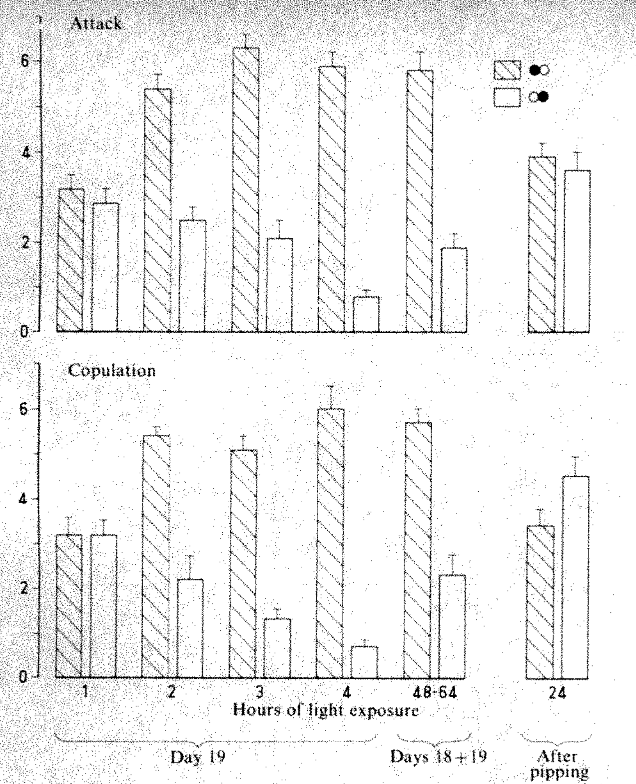


Fig. 4 Attack and copulation scores are given as overall means (±s.e.) calculated from scores recorded daily from days 6 to 11 of life. The symbols are as in Figs 1 and 2. Apart from the hours of light exposure indicated, the eggs were incubated in darkness. There is a significant elevation ($P<0.002$) in both scores for the left-treated groups (white columns) compared with the right-treated groups (black columns) after 2, 3, 4 and 48–64 h light exposure on day 19 of incubation, but not after 1 h on the same day or ~24 h from the time of pipping to hatching ($n=8$–12 for each group).

which suggests a bimodal distribution. This was also the case for chicks from eggs incubated similarly and treated in the right hemisphere. Further sampling will be necessary to confirm this (J. V. Zappia and L.J.R., in preparation), but it is possible that in the absence of light exposure before hatching individual chickens still have lateralized forebrain function, but that half are asymmetrical in one direction and half in the other direction, similar to pawedness in mice[14]. Light exposure during embryonic development may not generate lateralization in individuals, but synchronize the direction of lateralization in the population such that differences emerge between groups treated with glutamate in the left or right hemisphere.

On about day 19 of incubation, when the visual pathways between the tectum and the Wulst area of the forebrain are becoming functional, asymmetrical light input may synchronize the direction of functional asymmetry in the population, perhaps by stimulating blood flow to the left hemisphere in response to neuronal activity[15] and increasing protein synthesis[16] or some other metabolic process. The direction of asymmetry for attack and copulation therefore seems to be a consequence of the orientation of the embryo in the egg.

Denenberg[17] has reported that asymmetry of hemispheric function in the rat is enhanced by the experience of handling in early life. In this case environmental experience influences the degree of asymmetry, differing from the chicken data in which light experience seems to affect the direction but not the degree of asymmetry. However, both examples illustrate that interaction between genotype and experience determines the presence or form of functional asymmetry in the brain.

Received 5 January; accepted 26 March 1982.

1. Denenberg, V. H. *Behav. Brain Sci.* **4**, 1–21 (1981).
2. Nottebohm, F. in *Lateralization in the Nervous System* (eds Harnad, S., Doty, R. W., Goldstein, L., Jaynes, J. & Kranthamer, G.) 23–44 (Academic, New York, 1977).
3. Rogers, L. J. & Anson, J. M. *Pharmac. Biochem. Behav.* **10**, 679–686 (1979).
4. Howard, K. J., Rogers, L. J. & Boura, A. L. A. *Brain Res.* **188**, 369–382 (1980).
5. Rogers, L. J. *Acta XVII congr. int. orn.* **1**, 653–659 (1980).
6. Hambley, J. W. & Rogers, L. J. *Neuroscience* **4**, 677–684 (1979).
7. Rogers, L. J. & Hambley, J. W., *Behav. Brain Res.* **4**, 1–18 (1982).
8. Levy, J. in *Hemisphere Function in the Human Brain* (eds Dimond, S. J. & Beaumont, J. G.) 121–177 (Elek, London, 1974).
9. Morgan, M., in *Lateralization in the Nervous System* (eds Harnad, S., Doty, R. W., Goldstein, L., Jaynes, J. & Krauthamer, G.) 173–194 (Academic, New York, 1977).
10. Oppenheim, R. W. in *Studies on the Development of Behaviour and the Nervous System* Vol. 1 (ed. Gottlieb, G.) 163–244 (Academic, New York, 1978).
11. Freeman, B. M. & Vince, M. A. *Development of the Avian Embryo* (Chapman & Hall, London, 1974).
12. Sedláček, J. *Adv. Psychobiol.* **1**, 129–170 (1972).
13. Daniel, W. W. *Applied Non-parametric Statistics* (Houghton Mifflin, Boston, 1978).
14. Collins, R. L. *J. Hered.* **60**, 117–119 (1969).
15. Bondy, S. C. & Harrington, M. E. *Science* **199**, 318–319 (1978).
16. Bateson, P. P. G., Horn, G. & Rose, S. P. R. *Brain Res.* **39**, 449–465 (1972).
17. Denenberg, V. H., Garbanati, J., Sherman, G., Yutzey, D. A. & Kaplan, R. *Science* **201**, 1150–1151 (1978).

Is oxytocin an ovarian hormone?

D. Claire Wathes & R. W. Swann

Department of Anatomy, The Medical School, University of Bristol, Bristol BS8 1TD, UK

Oxytocin is a nonapeptide hormone produced by the hypothalamus and released from the posterior pituitary. It is cleaved from a large molecular weight precursor which is synthesized in cell bodies of hypothalamic magnocellular neurones and then packaged into membrane-bound granules[1]. During axonal transport the precursor is cleaved to produce oxytocin and an oxytocin-related neurophysin. Apart from its established roles in lactation and labour, oxytocin is thought to be an important regulator of the oestrous cycle; it causes luteolysis in cattle[2] and immunization against oxytocin increases the length of the ovine oestrous cycle[3]. McCracken[4] has proposed that oxytocin acts by stimulating prostaglandin $F_{2\alpha}$ ($PGF_{2\alpha}$) release from the uterine endometrium and that steroid hormones regulate $PGF_{2\alpha}$ secretion by altering the availability of endometrial oxytocin receptors[4,5]. Steriods may also have a direct influence on oxytocin release from the posterior pituitary, oestradiol-17β increasing and progesterone inhibiting oxytocin secretion in response to vaginal distension in the goat[6,7]. However, recent radioimmunoassay (RIA) measurements of peripheral oxytocin levels during the ovine oestrous cycle[8–10] consistently show that oxytocin and progesterone concentrations increase and decrease in synchrony during the luteal phase, reaching basal levels after luteolysis at a time when oestradiol-17β titres are known to rise[11]. Therefore, there may exist different controls of reflex release and basal secretion of oxytocin, and our discovery of large amounts of oxytocin within the ovine corpus luteum reported here may provide an explanation.

Peptides were extracted from ovine corpora lutea using the method developed by Walsh and Niall[12] for the isolation of porcine relaxin. The crude extract was applied to a Sephadex G-50 column. While testing the fractions from this column for relaxin-like activity in a rat uterine strip bioassay, it became apparent that some fractions were stimulatory rather than inhibitory. A similar finding has been reported previously in the cow[13]. The fractions were therefore measured by RIA for cross-reactivity with oxytocin (Fig. 1). A single peak of oxytocin-like immunoreactivity eluted between 330 and 400 ml in a position corresponding to that of synthetic oxytocin (Syntocinon. Sandoz). Dilution curves of material from this peak (henceforth referred to as luteal extract) showed parallelism with the oxytocin standard curve (Fig. 2). HPLC analysis of the extract revealed oxytocin immunoreactivity eluting in a position identical to that of synthetic oxytocin (Fig. 3).

Luteal extract was tested for oxytocin in two bioassays using concentrations calculated from the RIA results to correspond to the oxytocin standards. In a rat milk ejection assay, close

quantitative and qualitative agreement was achieved between the response to both luteal extract and oxytocin standard. In the rat uterine strip bioassay the estimated oxytocin-like activity approximated to one third of that measured by RIA but this discrepancy may have been due to the presence of a factor which inhibited uterine activity in the same Sephadex G-50 extract.

These data clearly indicate that the peptide extracted from the corpora lutea is authentic oxytocin. The yield of oxytocin from the ovaries of non-pregnant sheep was 2.6 μg per g of luteal tissue whereas the yield from pregnant sheep was only 34 ng per g. As the average weight of ovine corpora lutea is ~0.5 g and because there are normally 1–3 corpora lutea per animal, a non-pregnant ewe in the luteal phase of the oestrous cycle has ~2.6 μg of ovarian oxytocin. Lederis[14] calculated that the posterior pituitary and hypothalamus of the sheep together contained 18.2 μg oxytocin. Thus ovarian oxytocin represents

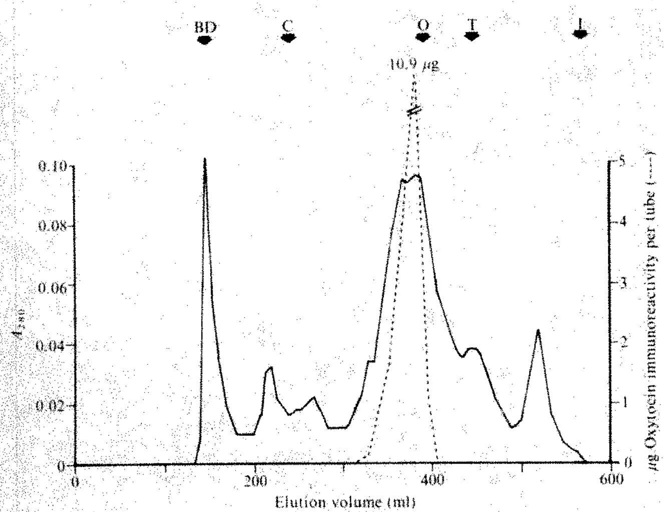


Fig. 1 Elution profile of a crude extract of 14.5 g ovine corpora lutea applied to a Sephadex G-50 (fine) column (120×2.4 cm) measured as absorbance at 280 nm. The eluant used was 0.1 M ammonium acetate *p*H 6.8 and the flow rate was 20.1 ml h^{-1}. Fractions (6.7 ml) were collected and lyophilized to remove ammonium acetate. The elution positions of various markers are shown: BD, blue dextran; C, cytochrome *c* (MW 12,300); O, oxytocin (MW 1,000); T, tyrosine (MW 181); I, ^{125}I. Paired fractions were assessed for oxytocin immunoreactivity by RIA. The rabbit antiserum used was raised against oxytocin coupled to thyroglobulin by the technique of Sofroniew *et al.*[24] and showed no significant cross-reaction with arginine-vasopressin, arginine-vasotocin or the rat neurophysins. ^{125}I-oxytocin, prepared by the chloramine-T method, was purified by chromatography on Sephadex G-25. The 4th International Standard for oxytocin (76/575) was used as a reference preparation.

~15% of the total hypothalamo-neurohypophysial store of hormone in the non-pregnant ewe during the luteal phase and ~0.2% during pregnancy.

Although it is possible that the presence of oxytocin in the ovary reflects its uptake from the blood, the high concentration of hormone suggests that it is synthesized there. However, proof of this hypothesis awaits the results of further experiments such as incorporation studies. If oxytocin is synthesized in a similar manner in the ovary and brain, one would also expect to find an ovarian neurophysin. We were unable to assay for ovine neurophysin but the Sephadex G-50 peak expected to contain molecules of molecular weight (MW) 10,000 (200–295 ml, Fig. 1) did contain small amounts of a component having the electrophoretic and HPLC characteristics of bovine neurophysin I.

Luteal oxytocin is probably not important for lactation in the ewe, as corpora lutea regress at parturition and a lactating ewe may not ovulate again for several months. Ovariectomized sheep can also deliver successfully, arguing against a role in parturition[15,16]. However, oxytocin measured in the peripheral

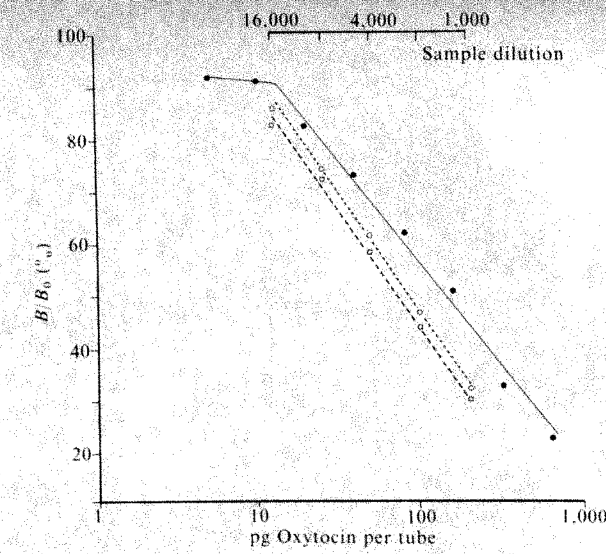


Fig. 2 Dilution curves for oxytocin standard (●) and two Sephadex-G50 fractions taken from separate extractions (○ and □). The ordinate shows bound ^{125}I-oxytocin (*B*) expressed as a percentage of ^{125}I-oxytocin bound in the absence of oxytocin (B_0). The slopes of the lines for the standards and samples did not differ significantly.

plasma during the oestrous cycle[8-10] could be of luteal origin as the increases and decreases in its concentration correspond to the waxing and waning of the corpus luteum. Additional evidence in support of this view is that ovariectomy leads to a decrease in circulating oxytocin[10] and that episodic release was not detected during seasonal anoestrus[9]. Circulating oxytocin concentrations also decline during pregnancy[8,9] corresponding to the much lower yield of oxytocin obtained from the ovaries of pregnant ewes in the present study.

The ovine corpus luteum is known to contain secretory granules, the release of which reaches a maximum in mid-cycle[17,18]. It has been suggested that these granules contain progesterone[17,19,20], but steroid secretion is not usually associated with granule formation and release. The present findings raise the possibility that these granules represent the ovarian

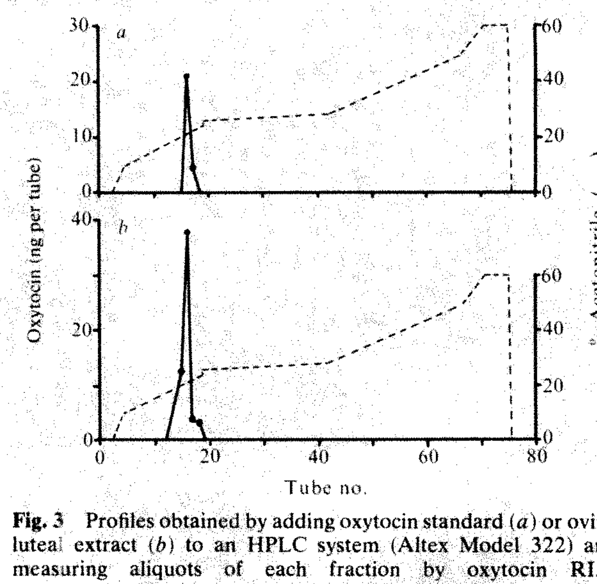


Fig. 3 Profiles obtained by adding oxytocin standard (*a*) or ovine luteal extract (*b*) to an HPLC system (Altex Model 322) and measuring aliquots of each fraction by oxytocin RIA. Chromatography was performed on a 5 mm (i.d.) × 10 cm stainless steel column packed in our laboratory with Hypersil ODS 5 μ. Samples were loaded and washed on to the column in 0.2 M Na_2HPO_4 *p*H 2.1 (adjusted with H_3PO_4) and eluted at a flow rate of 1 ml min^{-1} with an increasing gradient of acetonitrile (HPLC grade S, Rathburn). Fractions (1.2 ml) were collected. Vasotocin and arginine-vasopressin elute in tubes 11 and 12, respectively, using this gradient system.

oxytocin stores. Indeed another peptide hormone, relaxin, is packaged by the luteal cells into similar granules during pregnancy in the pig and rat[21,22].

The discovery of oxytocin in the corpus luteum raises a number of questions. If oxytocin were principally involved in the control of $PGF_{2\alpha}$ release at luteolysis, one would expect its concentration to peak shortly before the normal rise in $PGF_{2\alpha}$, but oxytocin levels decrease at about this time[8]. McCracken[4] has proposed that a rise in uterine oxytocin concentrations induced by oestradiol-17β is the first step in luteolysis but oestrogen levels similarly do not rise until progesterone concentrations have started to decline[23]. Thus the role of oxytocin during the oestrous cycle remains undetermined.

We thank Mr R. Jones, Miss J. Creek, Miss A. Moffatt and Mr P. Rees for technical assistance; S. D. Birkett for the radioimmunoassay; Dr J. A. Wakerley for the milk ejection experiment; Dr C. M. Wathes for the statistical analysis and Professors B. T. Pickering and D. G. Porter for their helpful discussions. The work was supported in part by grants from the ARC (AG 7/103), MRC (G8011540SA, G979/479) and the Wellcome Trust (9267/1.4P).

Note added in proof:
HPLC and RIA studies using bovine corpora lutea have shown that both oxytocin and bovine neurophysin I are present in the ovary in this species too.

Received 16 February; accepted 24 March 1982.

1. Pickering, B. T. *Essays Biochem.* **14,** 45–81 (1978).
2. Armstrong, D. T. & Hansel, W. *J. Dairy Sci.* **42,** 533–542 (1959).
3. Sheldrick, E. L., Mitchell, M. D. & Flint, A. P. F. *J. Reprod. Fert.* **59,** 37–42 (1980).
4. McCracken, J. A. *Adv. Prostaglandin Thromb. Res.* **8,** 1329–1344 (1980).
5. Roberts, J. S., McCracken, J. A., Gavagan, J. E. & Soloff, M. S. *Endocrinology* **99,** 1107–1114 (1976).
6. Roberts, J. S. *Endocrinology* **89,** 1029–1033 (1971).
7. Roberts, J. S. *Endocrinology* **89,** 1137–1141 (1971).
8. Webb, R., Mitchell, M. D., Falconer, J. & Robinson, J. S. *Prostaglandins* **22,** 443–453 (1981).
9. Sheldrick, E. L. & Flint, A. P. F. *Prostaglandins* **22,** 631–636 (1981).
10. Schams, D., Lahlou-Kassi, A. & Glatzel, P. *J. Endocr.* **92,** 9–13 (1982).
11. Baird, D. T. & Scaramuzzi, R. J. *J. Endocr.* **70,** 237–245 (1976).
12. Walsh, J. R. & Niall, H. D. *Endocrinology* **107,** 1258–1260 (1980).
13. Fields, M. J., Fields, P. A., Castro-Hernandez, A. & Larkin, L. H. *Endocrinology* **107,** 869–876 (1980).
14. Lederis, K. *Gen. comp. Endocr.* **1,** 80–89 (1961).
15. Neher, G. M. & Zarrow, M. X. *J. Endocr.* **11,** 323–330 (1954).
16. Fylling, P. *Acta. endocr.* **65,** 273–283 (1970).
17. Gemmel, R. T., Stacey, B. D. & Thorburn, G. D. *Biol. Reprod.* **11,** 447–462 (1974).
18. Paavola, L. G. & Christensen, A. K. *Biol. Reprod.* **25,** 203–215 (1981).
19. Gemmel, R. T. & Stacy, B. D. *Am. J. Anat.* **155,** 1–14 (1979).
20. Quirk, S. J., Wilcox, D. L., Parry, D. M. & Thorburn, G. D. *Biol. Reprod.* **20,** 1133–1145 (1979).
21. Kendall, J. Z., Plopper, C. G. & Bryant-Greenwood, G. D. *Biol. Reprod.* **18,** 94–98 (1978).
22. Anderson, M. L. & Long, J. A. *Biol. Reprod.* **18,** 110–117 (1978).
23. Baird, D. T., Land, R. B., Scaramuzzi, R. J. & Wheeler, A. G. *J. Endocr.* **69,** 275–286 (1976).
24. Sofroniew, M. V., Madler, M., Müller, O. A. & Scriba, P. C. *Z. analyt. Chem.* **290,** 163 (1978).

Noradrenergic modulation of dendrodendritic inhibition in the olfactory bulb

C. E. Jahr* & R. A. Nicoll

Departments of Pharmacology and Physiology,
University of California, San Francisco, California 94143, USA

Noradrenaline containing fibres in the central nervous system (CNS) have been characterized in detail by histochemical techniques[1–3]. A particularly striking feature of this neurotransmitter system is its diffuse pattern of innervation. Although noradrenaline generally inhibits spontaneous activity of neurones[4–7], in a number of systems[5,8–10] it can actually enhance orthodromic excitatory responses. Enhancement of excitatory transmission through sensory relay neurones is consistent with the view that noradrenergic systems are intimately involved in such behaviours as arousal and attention[11]. However, the basis for this enhancement is unclear. Two proposed hypotheses are a facilitation of excitatory transmitter action[5] and a disinhibitory action in which noradrenaline inhibits inhibitory interneurones[8]. The olfactory bulb is ideally suited for studying the cellular action of noradrenaline because the neuronal circuitry is relatively simple. Using intracellular recording techniques we have now found that noradrenaline attenuates the inhibitory feedback onto relay neurones, thus facilitating their firing, while glutamate and aspartate augment the inhibitory feedback.

The olfactory bulb has two major cell types: mitral cells, which relay signals to the cortex and granule cells, which are inhibitory interneurones. The interaction between these two types of cells involves a reciprocal dendrodendritic inhibitory pathway[12–18]. Briefly, impulses in mitral cell dendrites release excitatory transmitter (probably aspartate[19]) onto the dendrites of granule cells which, in turn, release the inhibitory transmitter, γ-aminobutyric acid (GABA) back onto mitral cells. In addition, the olfactory bulb receives a massive input of centrifugal fibres[20], some of which contain noradrenaline[21]. As some of the centrifugal fibre synapses contain spherical vesicles similar to those found in mitral cells, the acidic amino acids glutamate and aspartate may also be neurotransmitters in centrifugal fibres. The observed close approximation of the centrifugal synapses to the reciprocal synapses suggests that they may act to change the gain of the reciprocal inhibitory pathway.

Standard intracellular recording techniques were used in the present study and are described in detail elsewhere[16,17]. Briefly, turtles (*Pseudemis scripta elegans*) were decapitated and the olfactory bulbs removed and hemisected. The superfusate was a modified Ringer solution buffered with Tris(hydroxymethyl)aminomethane to *p*H 7.4 and bubbled with 100% O_2. Experiments were performed at 19 °C. The olfactory nerves and mitral cell axons were stimulated with either concentric bipolar or bipolar needle electrodes. Drugs were applied in the superfusate or by iontophoresis. The placement of the tip of

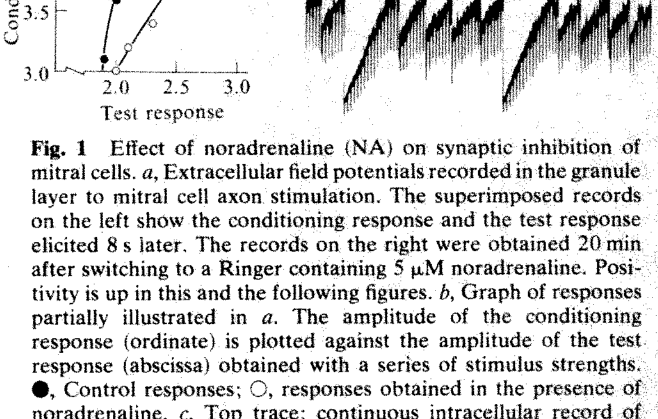


Fig. 1 Effect of noradrenaline (NA) on synaptic inhibition of mitral cells. *a*, Extracellular field potentials recorded in the granule layer to mitral cell axon stimulation. The superimposed records on the left show the conditioning response and the test response elicited 8 s later. The records on the right were obtained 20 min after switching to a Ringer containing 5 μM noradrenaline. Positivity is up in this and the following figures. *b*, Graph of responses partially illustrated in *a*. The amplitude of the conditioning response (ordinate) is plotted against the amplitude of the test response (abscissa) obtained with a series of stimulus strengths. ●, Control responses; ○, responses obtained in the presence of noradrenaline. *c*, Top trace: continuous intracellular record of orthodromic (arrows) and spontaneous i.p.s.ps in control superfusate. Lower trace; the same, 12 min after the addition of 50 μM noradrenaline. The downward deflections are responses of the membrane to constant current hyperpolarizing current pulses. The threshold for this cell was −50 mV.

* Present address: Department of Neurobiology, Harvard Medical School, 25 Shattuck Street, Boston, Massachusetts 02115, USA.

0028-0836/82/200227—03$01.00

the iontophoretic pipette in the external plexiform layer was facilitated by monitoring the focal extracellular field potentials with this pipette while it was being advanced through the tissue. The pipettes contained 1 M GABA (pH 4), 1 M sodium glutamate (pH 8) or 1 M sodium aspartate (pH 8).

Stimulation of the olfactory nerves or the mitral cell axons results in a complex of field potentials that can be recorded in the olfactory bulb. The largest wave recorded in the granule cell layer represents the excitation of granule cells[12-14]. Such stimulation results in a prolonged inhibition of the same wave evoked by a second identical stimulus. This paired pulse inhibition is an extracellular measure of the inhibitory postsynaptic potentials (i.p.s.ps) produced in a population of mitral cells that result from the activation of the dendrodendritic reciprocal synaptic pathway[13,14]. In all five experiments the addition of noradrenaline to the superfusate decreased this paired pulse inhibition (Fig. 1*a*), indicating that activation of the dendrodendritic inhibitory pathway was, at some point, decreased. This effect was more pronounced at moderate and high stimulus strengths (Fig. 1*b*). Noradrenaline also depressed the amplitude of the field potentials and therefore, in Fig. 1*a*, the stimulus strength was increased to evoke a conditioning response identical to that in the control.

It has been demonstrated with intracellular recording from mitral cells that the i.p.s.ps are GABA mediated, chloride dependent[16-18] and can be evoked by a single action potential in a mitral cell[16,17]. In the present experiments in control conditions, a number of mitral cells displayed successive spontaneous action potentials with intervening i.p.s.ps (Fig. 1*c*, top trace) which controlled the spontaneous action potential frequency. Thus, if the preparation was not stimulated, spikes occurred at a very regular interval which was determined by the amplitude and rate of decay of the i.p.s.ps. When the superfusate was switched to one containing noradrenaline (5–100 μM), both the i.p.s.p. following the spontaneous action potentials and the olfactory nerve-evoked i.p.s.ps (marked with arrows) were invariably shortened (Fig. 1*c*, bottom trace; $n = 11$). In addition, the peak conductance increase associated with the i.p.s.ps was decreased. The i.p.s.ps following spontaneous action potentials were also decreased in amplitude. In those cells which were not spontaneously firing action potentials noradrenaline had no effect on the membrane potential or conductance, although in some cells the threshold for spike initiation increased. This last effect might account in part for the depressant action of noradrenaline on the field potentials (see above) and the inhibitory effects reported *in vivo*[22].

The diminution of the i.p.s.p. could be due to the action of noradrenaline on either mitral or granule cells or both. As neither resting membrane potential nor membrane resistance of mitral cells was affected by nonadrenaline, its site of action might be on the granule cells, although it could exert its effects

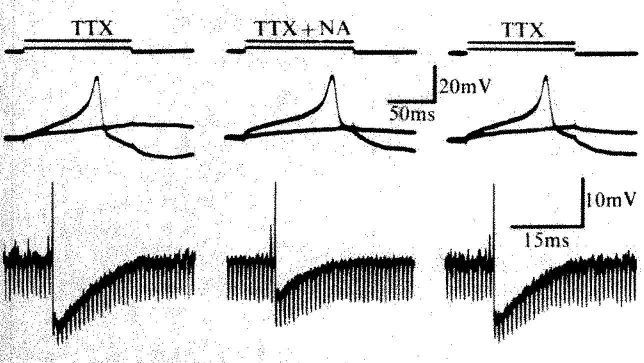


Fig. 2 Diminution of the i.p.s.p. in the presence of tetrodotoxin by noradrenaline. Top row: oscilloscope records of calcium spikes in tetrodotoxin (1 μM), after 15 min of superfusion with 100 μM noradrenaline, and after 23 min of washing in tetrodotoxin. Bottom row: simultaneous pen records of i.p.s.ps following calcium spikes. Resting potential was −60 mV throughout. The voltage calibration bar represent 1 nA for the current trace.

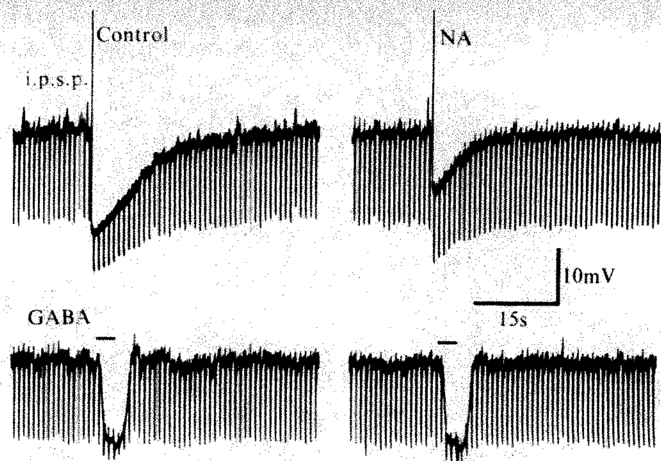


Fig. 3 Noradrenaline is not a GABA antagonist. The records on the left are the control i.p.s.p. (top) and control GABA response (bottom). The records on the right were obtained 24 min after the addition of noradrenaline (20 μM) to the superfusate. GABA was applied with a current of 80 nA. Resting potential was −54 mV throughout.

by blocking calcium entry (see refs 23, 24) into mitral cells. To examine this question further, the effects of noradrenaline were tested on mitral cell responses in the presence of tetrodotoxin. The calcium spikes that can be elicited in the presence of tetrodotoxin[15-17] were not reduced by concentrations of noradrenaline up to 100 μM while the i.p.s.p. was reduced substantially (Fig. 2; $n = 3$). Furthermore, this could be observed in cells in which spike threshold was unaltered.

Next, by iontophoresing GABA onto mitral cells we tested the possibility that noradrenaline might attenuate i.p.s.ps by blocking the action of the inhibitory transmitter. However, concentrations of noradrenaline that markedly reduced the size of the i.p.s.p. did not affect the iontophoretic GABA response (Fig. 3; $n = 4$). Thus, we conclude that the primary site of action of noradrenaline is on the granule cells. It has been reported that noradrenaline can enhance the action of GABA in the cerebellum[5]. We did not see such an enhancement, although small increases could have gone undetected. However, as shown above the overall effect of noradrenaline on synaptic inhibition in the olfactory bulb is depression.

In contrast to the actions of noradrenaline, iontophoretic glutamate enhanced the amplitude and duration of i.p.s.ps evoked by spontaneous action potentials (Fig. 4*a*; $n = 3$). This prolongation decreased the frequency of spikes without a change in spike threshold. When the iontophoretic current was increased, complete cessation of spontaneous action potentials occurred concomitant with a hyperpolarization of the membrane potential and a decrease in membrane resistence (Fig. 4*b*). Very short but intense ejections of glutamate resulted in fast hyperpolarizations associated with decreases in membrane resistance. Both effects outlasted the iontophoretic pulse and took the same form as i.p.s.ps (Fig. 4*c*). This hyperpolarizing action of glutamate could be observed in the presence of tetrodotoxin (Fig. 4*d*). However, after the addition of the GABA antagonist, bicuculline (Fig. 4*d*), or after blockade of synaptic transmission by cobalt, glutamate evoked a depolarization ($n = 8$). This indicates that the direct action of glutamate on mitral cells is a depolarization and suggests that, like the action of noradrenaline, glutamate acts on mitral cells primarily by modulating the release of GABA from granule cells.

Our observations suggest that noradrenaline and also enkephalin[25] decrease the i.p.s.p. recorded in mitral cells by acting on granule cells, causing them to release less GABA in response to the same stimulus. Glutamate and aspartate exert the opposite effect on granule cells. Based on extracellular recording techniques, McLennan[26] suggested that glutamate and noradrenaline might be acting in a similar way on granule

cells. The reason for this discrepancy is unclear. Noradrenaline may produce its disinhibitory effect by shunting or hyperpolarizing the granule cell membrane, thus attenuating excitatory input, or by interfering with excitation–release coupling[23,27]. Although a direct effect on the release process in mitral cells has not been completely ruled out, mitral cell calcium spikes are unaffected by noradrenaline, and thus any interference with release would presumably have to occur after calcium entry.

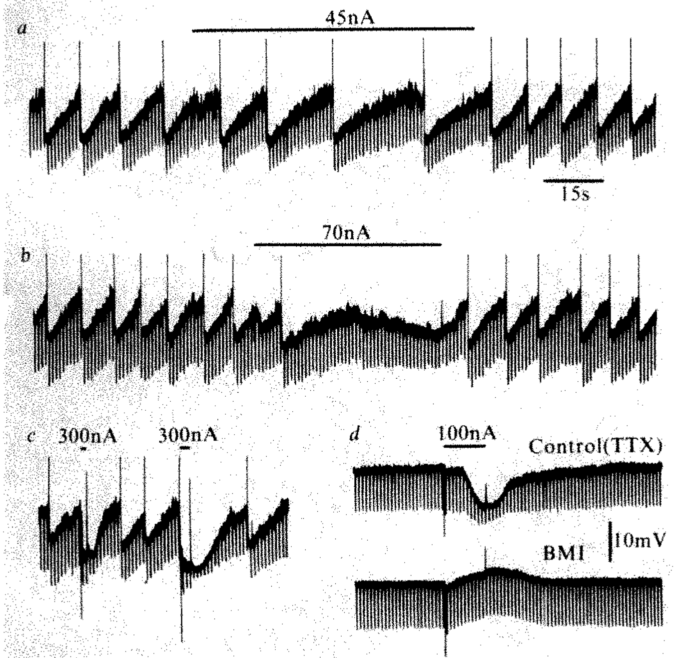


Fig. 4 Effect of glutamate on mitral cell i.p.s.ps and membrane potential. *a–c*, Iontophoretic applications of glutamate onto the same cell. Spike threshold was −57 mV throughout. *d*, Glutamate iontophoresis in a preparation perfused with tetrodotoxin (1 μM) before (top trace) and 20 min after the addition of 0.1 mM bicuculline methiodide (BMI) (bottom trace). Resting potential was −55 mV.

The hyperpolarizing action of glutamate and aspartate on mitral cells has been shown to result from the release of GABA from granule cells. This is presumably due to a depolarizing action of the amino acids on granule cells. The fact that doses of excitatory amino acid below that which caused the release of GABA still enhanced the evoked release of GABA suggests that subthreshold depolarization of granule cells facilitates excitation–secretion coupling. This result agrees with recent findings in invertebrate neurones[28].

In conclusion, it is proposed that noradrenaline, enkephalin and glutamate alter mitral cell excitability primarily by indirect actions; noradrenaline and enkephalin block GABA release from granule cells, whereas glutamate enhances GABA release. The granule cell, then, is the final common pathway on which the centrifugal fibres and their transmitter candidates exert presynaptic control and thereby modulate the flow of information through the olfactory bulb. Such a disinhibitory role for noradrenaline in a sensory pathway fits well into the proposed role of the locus coeruleus noradrenergic system in arousal[11], a behavioural state associated with enhanced neuronal responsiveness[29].

This research was supported by NIH grant NS 16485, RCDA NS 00287 to R.A.N., and by the Klingenstein Fund.

Received 1 October 1981; accepted 12 March 1982.

1. Moore, R. Y. & Bloom, F. E. *A. Rev. Neurosci.* **2**, 113–168 (1969).
2. Fallon, J. G. & Moore, R. Y. *J. comp. Neurol.* **180**, 533–544 (1978).
3. Morrison, J. H., Molliver, M. E., Grzanna, R. & Coyle, J. T. *Neuroscience* **6**, 139–158 (1981).
4. Bloom, F. E. in *Psychopharmacology—A 20 Year Progress Report* (eds Lipton, M. E., Killam, K. C. & Di Mascio, A.) 131–142 (Raven, New York, 1978).
5. Woodward, D. J., Moises, H. C., Waterhouse, B. D., Hoffer, B. J. & Freedman, R. *Fedn Proc.* **38**, 2109–2116 (1979).
6. Segal, M. *Brain Res.* **206**, 107–128 (1981).
7. Langmoen, I. A., Segal, M. & Andersen, P. *Brain Res.* **208**, 349–362 (1981).
8. Nakai, Y. & Takaori, S. *Brain Res.* **71**, 47–60 (1974).
9. Rogawski, M. A. & Aghajanian, G. K. *Nature* **287**, 731–734 (1980).
10. Mueller, A. L., Hoffer, B. J. & Dunwiddie, T. V. *Brain Res.* **214**, 113–126 (1981).
11. Aston-Jones, G. & Bloom, F. E. *J. Neurosci.* **1**, 887–900 (1981).
12. Rall, W. & Shepherd, G. M. *J. Neurophysiol.* **31**, 884–915 (1968).
13. Nicoll, R. A. *Brain Res.* **14**, 157–172 (1969).
14. Mori, K. & Takagi, S. F. *J. Physiol., Lond.* **305**, 171–196 (1980).
15. Mori, K., Nowycky, M. C. & Shepherd, G. M. *J. Physiol., Lond.* **314**, 295–309 (1981).
16. Jahr, C. E. & Nicoll, R. A. *Science* **207**, 1473–1475 (1980).
17. Jahr, C. E. & Nicoll, R. A. *J. Physiol., Lond.* (in the press).
18. Nowycky, M. C., Mori, K. & Shepherd, G. M. *J. Neurophysiol.* **46**, 639–648 (1981).
19. Collins, G. G. S. & Probett, G. A. *Brain Res.* **209**, 231–234 (1981).
20. Price, J. L. & Powell, T. P. S. *J. Cell Sci.* **7**, 157–187 (1970).
21. Halasz, N., Ljungdahl, A. & Hokfelt, T. *Brain Res.* **154**, 253–271 (1978).
22. Bloom, F. E., Costa, E. & Salmoiraghi, G. C. *J. Pharm. exp. Ther.* **146**, 16–23 (1964).
23. Dunlap, K. & Fischbach, G. D. *J. Physiol., Lond.* **317**, 519–535 (1981).
24. Horn, J. P. & MacAfee, D. A. *J. Physiol., Lond.* **301**, 191–204 (1980).
25. Nicoll, R. A., Alger, B. E. & Jahr, C. E. *Nature* **287**, 22–25 (1980).
26. McLennan, H. *Brain Res.* **29**, 177–184 (1971).
27. Shapiro, E., Castellucci, V. F. & Kandel, E. R. *Proc. natn Acad. Sci. U.S.A.* **77**, 1185–1189 (1980).
28. Shapiro, E., Castellucci, V. F. & Kandel, E. R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 629–633 (1980).
29. Livingstone, M. S. & Hubel, D. H. *Nature* **291**, 554–561 (1981).

Differential release of serotonin and histamine from mast cells

Theoharis C. Theoharides, Philip K. Bondy*, Nikolaos D. Tsakalos† & Philip W. Askenase

Department of Internal Medicine, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510, USA and
* Veterans Administration Medical Center, West Spring Street, West Haven, Connecticut 06516, USA

Exocytosis dependent on calcium and metabolic energy has been established as the mechanism for the release of membrane-bound secretory products from various exocrine, endocrine and neural cells[1]. This has also been shown to be the case in mast cells, which have been used increasingly as a model secretory system[2]. The secretory granules of mast cells contain several mediators[3], some of which, such as histamine, are known to participate in many immune reactions and allergic diseases[4,5]. Because of mast cell involvement in these clinical syndromes, as well as the role of histamine in gastric acid secretion[6] and possibly in brain pathophysiology[7], there has been great interest in the pharmacological modulation of histamine release from mast cells[8]. Serotonin is also stored in mast cell granules of several species but much less is known about its secretion. Because histamine and serotonin may have divergent functions in delayed hypersensitivity[4,9], we hypothesized that these amines could undergo differential release. We now report that the tricyclic antidepressant drug amitriptyline (Elavil) inhibits histamine release from stimulated mast cells while permitting the release of serotonin. In these conditions, exocytosis of secretory granules is largely prevented, but serotonin is released by an unknown process which still requires calcium and metabolic energy. The ability to secrete differentially expands the physiological potential of the mast cell, and suggests that release of serotonin may not always indicate mast cell secretion via exocytosis of secretory granules.

Rat peritoneal mast cells were collected, purified (90% purity) and incubated as described previously[10]. Mast cells that had been preincubated with ^{3}H-serotonin to load them with this releasable amine[11] and then incubated with amitriptyline (10^{-7}–10^{-4} M) for 5 min showed depressed histamine release

† Permanent address: Department of Internal Medicine, Aristotelian University School of Medicine, Thessaloniki, Greece.

0028–0836/82/200229—03$01.00

Table 1 Amitriptyline permits differential release from mast cells

Amitriptyline (M)	Mast cell release (% total): Histamine 48/80	Histamine IgE+Ag	Serotonin 48/80	Serotonin IgE+Ag	Ruthenium red staining 48/80	Ruthenium red staining IgE+Ag
—	71±3	40±9	83±4	46±6	83±6	52±1
10^{-7}	70±4	38±7	83±5	45±8	82±7	50±5
10^{-6}	65±5	30±8	78±6	38±6	79±5	41±6
5×10^{-6}	59±6	21±5	75±3	37±7	68±6	35±4
10^{-5}	40±3	17±6	74±4	39±5	57±4	28±3
5×10^{-5}	27±3	12±4	72±3	34±3	34±3	13±3
10^{-4}	11±2	7±4	76±5	39±2	19±3	9±1
	Inhibition of secretion (% total)					
10^{-4}	85	83	8	15	77	83

Effect of amitriptyline on mast cell release of histamine and serotonin. Mast cells (90% purity, 10^6 cells ml^{-1}) were first incubated in HEPES-buffered Locke's solution[10] with 5 μCi ml^{-1} ^3H-serotonin (29.8 Ci $mmol^{-1}$; NEN) for 1 h at 37 °C, and were then incubated (10^5 cells per tube) with 0–10^{-4} M amitriptyline for 5 min, followed by an additional 5 min incubation with compound 48/80 (1 μg ml^{-1}) or 30 min incubation with anti-ovalbumin hybridoma IgE antibody (1:1,000 dilution) and ovalbumin (Ag, 100 μg ml^{-1}) in the presence of phosphatidylserine (100 μM). The histamine released was assayed fluorometrically[38] while the serotonin released was measured by liquid scintillation counting[11], both in duplicate portions. Ruthenium red (0.005%) staining of mast cells was performed as described previously[14]. Spontaneous release was 1.8% for serotonin, 3.6% for histamine. Results represent the mean ± s.d. of five separate determinations.

but nearly unaltered serotonin release when stimulated either nonspecifically with the classic mast cell secretagogue[12] compound 48/80 (1 μg ml^{-1}) or immunologically by addition of 1:1,000 dilution of mouse ascites fluid from an IgE antibody-secreting hybridoma[13] and antigen (ovalbumin, 100 μg ml^{-1}) in the presence of 100 μM phosphatidylserine (Table 1). When experiments were performed with unpurified peritoneal cells (10% mast cells), higher concentrations of amitriptyline were required to permit differential release. Exocytosis of secretory granules was identified by selective staining with ruthenium red which, by virtue of its relative inability to cross the plasma membrane and its high affinity for proteoglycans of the secretory granule core, binds to and stains only those granules exposed to the extracellular fluid[14]. This assay showed that depression of histamine release by amitriptyline was accompanied by a substantial reduction in apparent granule exocytosis (Table 1). To examine the requirements necessary for the release of serotonin but not histamine, mast cells were preincubated with either EDTA (5×10^{-3} M for 2 h) or antimycin A (10^{-6} M) and 2-deoxyglucose (5×10^{-4} M) for 30 min followed by amitriptyline and then compound 48/80, or IgE and antigen as before. With calcium depletion or deprivation of metabolic energy, serotonin failed to appear in the supernatant (Table 2). Thus, differential release of serotonin seemed to be calcium and energy dependent.

Although histamine release has been the traditional measure of mast cell secretion, release of radiolabelled serotonin—presumably incorporated into mast cell granules from the extracellular medium—has been used increasingly as an alternative means of quantifying mast cell secretion because of its simplicity and sensitivity, and because it has so far been shown to parallel the release of histamine[11,15–17]. However, our results indicate that serotonin can be released in the absence of substantial histamine secretion, and in the absence of demonstrable exocytosis of secretory granules. To exclude the possibility that differential release was solely due to an unknown ability of exogenous radiolabelled serotonin to bind to some location from which it could be displaced without exocytosis of secretory granules, we used a HPLC system to assay endogenous (preformed and granule-stored) serotonin both in supernatants and in cell pellets of mast cells stimulated to secrete by compound 48/80 in the presence of amitriptyline (10^{-4} M). In conditions where there was only ~11% release of histamine and 19% staining by ruthenium red (Table 2), there was ~67% release of endogenous serotonin. These results show that activation of mast cells in the presence of amitriptyline resulted in similar release of both exogenous and endogenous serotonin.

Data from three previous studies suggest that serotonin may indeed be secreted differentially from mast cells[18–20]. Other studies have reported that different mast cell secretory products may be packaged either in distinct granules or in other cytoplasmic components from which they may be mobilized separately[21–22]. However, as serotonin and histamine are both thought to be stored in the secretory granules of mast cells, it is necessary to postulate a mechanism whereby serotonin, and not histamine, could be transported from the granules to the cell exterior. It is possible that in some circumstances serotonin may be preferentially moved from the secretory granules to another compartment, perhaps in small cytoplasmic vesicles from which it may then be secreted by energy- and calcium-requiring exocytosis. There is some evidence that serotonin is present in the cytoplasm of mast cells[23,24], and a transport system for serotonin that involves a specific serotonin binding protein released by exocytosis together with serotonin has been demonstrated in the nervous system[25]. Furthermore, a vesicular transport mechanism has recently been identified in guinea pig basophils[26], and we have identified proteins in rat basophil leukaemia cells and mast cells that bind serotonin[27]. We suggest that specific serotonin-binding proteins, perhaps in cytoplasmic vesicles, may be involved in differential release by mast cells.

The ability of amitriptyline to inhibit histamine release (which has also been shown in human basophils[28]) and to permit differential release of serotonin, may be partly related to its ability to inhibit phosphatidylinositol turnover[29], which has been linked to the cascade of events involved in mast cell secretion[30]. This differential release contrasts with the explosive release of many mast cell granule constituents that is commonly seen in allergic and immediate hypersensitivity disease mechanisms, and is often used to raise doubts about the physiological usefulness of such a dramatic reaction[31,32].

Our present finding, that there may be a way of triggering the selective release of mast cell mediators, is also supported by recent *in vivo* work. In these experiments, immunized mice were challenged with antigen and developed skin swelling responses associated with mast cell degranulation, release of serotonin and formation of gaps between endothelial cells[33]. However, mice treated with pargyline—another drug permitting differential secretion from mast cells[34]—showed no mast cell degranulation but had skin swelling and evidence of

Table 2 Energy and calcium dependence of differential release of serotonin

Stimulus	No inhibitory drug: Histamine —	Histamine No Ca^{2+}	Histamine No energy	Serotonin —	Serotonin No Ca^{2+}	Serotonin No energy	Amitriptyline (10^{-4}M): Histamine —	Histamine No Ca^{2+}	Histamine No energy	Serotonin —	Serotonin No Ca^{2+}	Serotonin No energy	Trypan blue staining —	Trypan blue staining No Ca^{2+}	Trypan blue staining No energy
48/80	70±3	7±2	6±2	81±5	5±2	8±3	12±5	4±1	7±2	81±7	9±2	8±3	2±1	3±1	5±1
IgE+Ag	38±9	6±3	5±1	36±8	9±3	7±2	9±6	5±2	4±1	42±6	7±2	5±2	1±1	2±1	3±1

Effect of energy or calcium deprivation on differential release of serotonin in the presence of amitriptyline (10^{-4}M) and in response to either compound 48/80 (1 μg ml^{-1}) or IgE antibody plus appropriate antigen as described in Table 1 legend. Mast cell viability was tested with the Trypan blue (0.04%) exclusion method and was >94%. Results represent the mean ±s.d. of five separate determinations.

serotonin-induced gaps between endothelial cells[35]. These results, with the recently demonstrated ability of mast cells to secrete individual granules in well localized areas of their plasma membrane[36,37], indicate that mast cells could participate in previously unrecognized functions, some of which may be involved in syndromes that are alleviated by antidepressant agents.

This work was supported in part by USPHS and NIH grants AI-12211, AI-11077 and AI-10497 and AI-17555.

Received 7 December 1981; accepted 2 March 1982.

1. Douglas, W. W. Ciba Fdn Symp. 54, 61–90 (1978).
2. Garland, L. G. & Mongar, J. L. Symp. Soc. exp. Biol. 30, 193–218 (1976).
3. Ho, P. C., Lewis, R. A., Austen, K. F. & Orange, R. P. Comprehensive Immun. 6, 179–228 (1979).
4. Askenase, P. W. Prog. Allergy 23, 199–320 (1977).
5. Kazimierczak, W. & Diamant, B. Prog. Allergy 24, 295–365 (1978).
6. Sell, A. & Walsh, J. H. A. Rev. Physiol. 41, 35–53 (1979).
7. Green, J. P., Maayani, S., Weinstein, H. & Hough, L. B. Psychopharmac. Bull. 16, 36–38 (1980).
8. Foreman, J. C. & Lichtenstein, L. M. A. Rev. Med. 31, 181–190 (1980).
9. Gershon, R. K., Askenase, P. W. & Gershon, M. D. J. exp. Med. 142, 732–747 (1975).
10. Theoharides, T. C. & Douglas, W. W. Endocrinology 102, 1637–1640 (1978).
11. Morrison, P. C., Roser, J. F., Henson, P. M. & Cochrane, C. G. J. Immun. 112, 573–582 (1975).
12. Röhlich, R., Anderson, P. & Uvnäs, B. J. Cell Biol. 51, 465–483 (1971).
13. Böttcher, J., Hammering, G. & Kapp, J. F. Nature 275, 761–762 (1978).
14. Nemeth, E. F. & Douglas, W. W. Naunyn-Schmiedebergs Archs Pharmak. 302, 153–163 (1978).
15. Otsuki, J. A., Grassick, R., Seymour, D. & Kind, L. S. Immun. Commun. 5, 27–39 (1976).
16. Mazinque, C., Dessaint, J.-P. & Capron, A. J. immun. Meth. 21, 65–77 (1978).
17. Chasin, M., Scott, C., Shaw, C. & Persico, F. Int. Archs Allergy appl. Immun. 58, 1–10 (1979).
18. Goldstein, D. J., Finkielman, S. & Nahmond, V. E. Medicine 34, 584–585 (1974).
19. Miller, P. & Church, M. K. Int. Archs Allergy appl. Immun. 52, 53–58 (1976).
20. Ichikawa, A., Kaneko, H., Mori, Y. & Tomita, K. Biochem. Pharmac. 26, 197–202 (1977).
21. Lynch, S. M., Austen, K. F. & Wasserman, S. F. J. Immun. 121, 1394–1399 (1978).
22. Sannes, P. S. & Spicer, S. S. Am. J. Path. 94, 447–456 (1979).
23. Calsson, S.-A. & Ritzen, M. Acta physiol. scand. 77, 449–464 (1969).
24. Gustafsson, B. Int. Archs Allergy appl. Immun. 63, 121–128 (1980).
25. Tamir, H. & Gershon, M. D. J. Neurochem. 33, 35–44 (1979).
26. Dvorak, A. M. et al. Lab. Invest. 42, 263–276 (1980).
27. Tamir, H., Theoharides, T. C., Gershon, M. D. & Askenase, P. W. J. Cell Biol. (in the press).
28. Lichtenstein, L. M. & Gillepsie, E. J. Pharmac. exp. Ther. 192, 441–450 (1975).
29. Smith, T. L. & Hauser, G. Biochem. Pharmac. 28, 1759–1763 (1979).
30. Cockroft, S. & Gomperts, B. D. Biochem. J. 178, 681–687 (1979).
31. Padawer, J. Am. J. Anat. 141, 299–302 (1974).
32. Kaliner, M. A. New Engl. J. Med. 501, 498–500 (1979).
33. Askenase, P. W., Bursztajn, S., Gershon, M. D. & Gershon, R. K. J. exp. Med. 152, 1358–1374 (1980).
34. Askenase, P. W. & Theoharides, T. C. Fedn Proc. 39, 905 (1980).
35. Askenase, P. W., Scwhartz, A., Siegel, J. N. & Gershon, R. K. Int. Archs Allergy appl. Immun. 66 (Suppl. 1) 225–233 (1981).
36. Lawson, D., Fewtrell, C. & Raff, M. C. J. Cell Biol. 79, 394–400 (1978).
37. Theoharides, T. C. & Douglas, W. W. Science 201, 1143–1145 (1978).
38. Kremzner, L. T. & Wilson, I. B. Biochim. biophys. Acta 50, 364–367 (1961).

H-2K-, *H-2I*- and *H-2D*-restricted hybridoma contact sensitivity effector cells

M. Minami, K. Okuda, M. E. Sunday & M. E. Dorf

Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA

Mice primed with 4-hydroxy-3-nitrophenylacetyl-*O*-succinimide (NP-*O*-Su) generate hapten-specific T cell-mediated contact or cutaneous hypersensitivity (CS) responses[1]. Analysis of the effector T cells mediating these *in vivo* responses suggested that multiple T-cell subpopulations were involved[2]. It has not been possible previously to separate these cell populations physically by conventional means. Therefore, to separate these cells and to provide a uniform source of CS effector cells for future analyses, we hybridized the CS effector cell population with AKR-derived BW5147 thymoma cells. We now describe five *H-2*-restricted hybridoma T-cell clones isolated from this fusion. Each hybridoma possesses hapten-specific CS activity and four also cross-react with 4-hydroxy-3-nitrophenylacetyl (NP) derivatives. Two of the hybridoma clones are specific for NP associated with *H-2I* gene products; two others for NP associated with *H-2D* products; and one for NP associated with *H-2K* gene products.

C57BL/6 mice were immunized with 2 mg NP-*O*-Su; 6 days later T cells were prepared from the regional lymph nodes as previously described for the enrichment of Ts_3 suppressor cells[3]. The hybridization procedure has been reported elsewhere[3], in fact the same series of hybridoma cells were used to screen for CS and suppressor cell activity. The hybridoma colonies were screened for CS reactivity by injecting 10^5 irradiated cells (2,000 R) together with either 40 μg of NP–bovine serum albumin (BSA) or 5×10^5 NP-coupled C57BL/6 spleen cells into the left footpad. Maximal antigen-specific swelling responses occurred 24–48 h after injection.

The specificity of five CS reactive hybridoma clones is demonstrated in Table 1. Two types of NP-reactive hybridoma cell lines were distinguished. The B6-CSI-15 and B6-CSI-78 lines demonstrated significant footpad swelling responses when given together with NP-coupled proteins (for example, NP–bovine γ-globulin, NP–BGG), but failed to respond when transferred with carrier protein (BGG) alone. Similar NP-specific swelling responses were noted when the hybridoma cells were injected with other NP-coupled proteins, including NP–BSA (Table 3) and NP–keyhole limpet haemocyanin (KLH) (data not shown). However, these hybridoma lines failed to give NP-specific responses when administered with NP-coupled syngeneic C57BL/6 spleen cells. In contrast, the B6-CSK-21, B6-CSD-35 and B6-CSD-52 hybrids demonstrated NP-specific swelling responses when transferred with NP-coupled C57BL/6 spleen cells, but failed to respond to trinitrophenyl (TNP)-coupled C57BL/6 spleen cells or NP-coupled proteins. Significant levels of footpad swelling were elicited after local injection of as few as 10^3 hybridoma cells plus antigen. Injections of 10^4 or 10^5 hybridoma cells resulted in larger swelling responses, while injection of 10^6 cells gave swelling responses only slightly larger than those induced by 10^5 cells. In contrast, when 10^3–10^6

Table 1 Specificity of hybridoma CS effector cells

Hybridoma cell line	Challenge antigen NP–BGG	NIP–BGG	BGG	NP-cells	NIP-cells	TNP-cells
BW5147	0.3±0.9	1.0±0.6	0.3±0.3	1.5±0.6	NT	NT
B6-CSI-15	8.3±1.7*	6.0±0.6*	0.6±0.3	0.8±0.6	NT	NT
B6-CSI-78	8.3±1.5*	7.3±1.5*	0.3±0.3	1.5±0.9	NT	NT
BW5147	0.9±1.3	NT	NT	0.3±0.7	0.3±0.3	1.3±0.7
B6-CSK-21	0.9±0.6	NT	NT	8.0±0.6*	10.0±0.6*	0.3±0.3
B6-CSD-35	0.9±0.9	NT	NT	8.3±0.3*	0.6±0.3	0.0±0.6
B6-CSD-52	1.3±0.6	NT	NT	8.3±0.3*	8.7±0.9*	1.3±0.9

1×10^5 irradiated BW5147 cells or hybridoma cells were mixed with 40 μg NP–BGG, NIP–BGG, BGG, or 5×10^5 NP-, NIP-, or TNP-conjugated C57BL/6 spleen cells, and these were then injected into the left footpad of C57BL/6 recipients in 25 μl Hanks' balanced salt solution. Footpad swelling, measured as the difference in footpad thickness between the left and right footpads 24 h after antigenic challenge, is recorded in units of 10^{-3} cm. The groups each contain 4–5 mice. NT, not tested.

* Significant swelling response ($P<0.01$).

0028–0836/82/200231—03$01.00

Table 2 Phenotypes of hybridoma CS effector cells

Hybridoma cell line	Antibody specificity Thy-1.1	Thy-1.2	H-2D^b	I-A^b	I-J^b	NP^b	CGAT	Mouse immuno-globulin
BW5147	98	7	3	5	0	2	0	1
B6-CSI-15	98	98	84	0	3	24	2	0
B6-CSI-78	98	98	74	3	2	4	3	0
B6-CSK-21	98	98	74	7	2	3	2	2
B6-CSD-35	100	98	73	6	2	2	1	0
B6-CSD-52	98	7	78	5	2	2	0	0
B6-Ts$_1$-3	98	5	84	0	53	31	0	0

Monoclonal hybridoma anti-Thy-1.1, -Thy-1.2 (NEN) and -I-A^b antibodies (Dr T. Springer, Sidney Farber Cancer Institute, Massachusetts), conventional anti-H-2D^b ((A × 18R)F$_1$ anti-B10) and anti-I-J^b (5R anti-3R) alloantisera, guinea pig anti-NP^b and anti-CGAT idiotypes, and polyvalent guinea pig anti-mouse immunoglobulin antisera (prepared as detailed elsewhere[6]) were used, together with predetermined optimal concentrations of rabbit complement, to lyse the hybridoma cells. The results are expressed as per cent specific lysis and represent pooled results from 2–6 experiments. Less than 15% lysis is considered insignificant. The B6-Ts$_1$-3 T-cell hybridoma line was used as a control. Characterization of monoclonal Ts$_1$ suppressor cells is detailed elsewhere[6]. CGAT is the common idiotype on anti-GAT antibodies[10].

hybridoma cells were given intravenously (i.v.) and the mice were challenged with antigen in the footpads, no swelling response was observed (data not shown). This may be due to the failure of the hybridoma cells to properly migrate to the site of antigenic challenge. However, note that the use of a local transfer system has limitations: as helper[4] and killer[4,5] T cells can also induce delayed inflammatory responses, it is difficult to determine whether these hybridoma cells are CS effector cells or represent hybridomas derived from distinct T-cell subsets. This may be a semantic problem[2].

Previous reports[1,2] have indicated that NP-specific CS effector cells derived from Igh^b-bearing C57BL/6 mice cross-reacted with NP derivatives such as 4-hydroxy-5-iodo-3-nitrophenylacetyl (NIP). To evaluate the fine specificity of the CS reactive hybridoma clones isolated, they were injected with either NIP-derivatized proteins or cells. As shown in Table 1, four of the five hybridomas demonstrated significant levels of NIP cross-reactivity. Previous analyses of the C57BL/6-derived NP-*O*-Su-induced CS effector cell population did not reveal a non-NIP-cross-reactive population[1,2] but these previous experiments involved adoptive transfers of a heterogeneous population of *H-2*-restricted cells. Thus the presence of some NIP-cross-reactive clones in the heterogeneous population probably prevented detection of other non-cross-reactive clones.

To further distinguish between these hybridoma clones, the cells were phenotyped by microcytoxicity testing[6] using antisera reactive with *Thy-1-*, *H-2-* and *Igh-V*-encoded determinants derived from the C57BL/6 parental strain. The AKR (*H-2^k*, *Igh^d*)-derived BW5147 thymoma cells and the NP-specific suppressor T-cell hybridoma line B6-Ts$_1$-3 of C57BL/6 (*H-2^b*, *Igh^b*) origin, were used as controls. All the lines carried the BW5147-derived Thy 1.1 marker (Table 2), as shown by the high percentage of lysis observed after treatment with monoclonal anti-Thy-1.1 antibody and complement. Most of the hybridoma clones also carried the Thy-1.2 marker derived from the spleen cell donors. The B6-CSD-52 and B6-Ts$_1$-3 lines were exceptions—this may be due to the loss of the chromosome carrying the *Thy-1.2* gene[7,8]. All the C57BL/6-derived hybridomas specifically reacted with anti-H-2D^b antisera. However, none of the hybridomas demonstrated significant reactivity with monoclonal anti-I-A^b or conventional anti-I-J^b antibodies[9]. The cytolytic activity and specificity of the anti-I-J^b antisera described previously[6] were confirmed by testing the I-J^b-bearing B6-Ts$_1$-3 hybridoma suppressor line.

To evaluate the nature of the antigen receptor on these hybridomas, the cells were tested for reactivity with guinea pig anti-NP^b idiotypic antisera[6]. Control guinea pig anti-idiotypic antisera directed against a different idiotypic determinant associated with anti-GAT (poly(Glu^{60}Ala^{30}Tyr^{10})) antibodies[10] and polyvalent guinea pig anti-immunoglobulin antisera were used as controls. Only the B6-CSI-15 and the control B6-Ts$_1$-3 hybridoma lines demonstrated specific lysis with anti-NP^b antisera. All cloned sublines of B6-CSI-15 also demonstrated <50% lysis with anti-NP^b antisera, suggesting that the idiotypic receptor is poorly expressed. Similar findings were reported for the idiotypic receptors on suppressor cell hybridomas[11]. In addition, note that only one of the four hybridoma lines which cross-reacted with NIP hapten (Table 1) carried detectable levels of NP^b-related idiotypic determinants, indicating that at the T-cell level, NIP cross-reactivity does not correlate with the presence of NP^b idiotypic determinants. These results are also consistent with data which indicate that B6-derived CS effector cells are generally not susceptible to complement-mediated lysis with anti-NP^b antisera (unpublished data).

The hybridoma CS effector lines were further distinguished by evaluating the genetic restrictions of activation. The B6-CSI-15 and B6-CSI-78 hybridoma lines only demonstrated NP-specific reactivity when transferred into I-A^b-bearing recipients (Table 3). Thus, transfer of B6-CSI-15 and B6-CSI-78 cells together with NP–BSA yielded significant footpad swelling responses in I-A^b-bearing C57BL/6 and B10.A(5R) recipients, but not in strains lacking an *I-A* subregion derived from the *H-2^b* haplotype. In contrast, when B6-CSK-21 cells were transferred together with NP-coupled syngeneic spleen cells, footpad

Table 3 Genetic restriction of hybridoma CS effector cells

Hybridoma cell line	Challenge antigen	Recipient strains B10.A(5R) *bbbkkddd*	B10.MBR *bkkkkkkq*	B10.A(4R) *kkbbbbbb*	B10.A(2R) *ddddddb*	B10.A *kkkkkddd*	C57BL/6 *bbbbbbbb*
BW5147	NP–BSA	1.0±0.4	1.0±0.4	1.0±0.7	NT	1.0±0.4	1.3±0.6
B6-CSI-15	NP–BSA	7.3±1.4*	1.0±0.9	0.3±1.3	NT	1.0±0.7	15.0±1.6*
B6-CSI-78	NP–BSA	9.8±1.4*	1.3±0.6	1.8±0.9	NT	1.5±0.6	10.5±1.6*
BW5147	NP-cells	0.8±0.5	0.8±0.3	NT	0.8±0.5	1.0±0.4	0.3±0.3
B6-CSK-21	NP-cells	10.8±2.3*	9.8±1.4*	NT	0.3±0.5	1.3±0.9	9.8±0.8*
B6-CSD-35	NP-cells	0.5±0.6	0.5±0.6	NT	10.3±0.9*	1.0±0.4	17.0±1.8*
B6-CSD-52	NP-cells	2.5±1.0	0.5±1.0	NT	8.0±1.1*	0.8±0.6	9.0±1.0*

See Table 1 legend for protocol. Recipients received 5×10^5 syngeneic hapten-modified spleen cells.

swelling responses were noted only in C57BL/6, B10.A(5R) and B10.MBR recipients. Thus, the activity of the B6-CSK-21 hybridoma cell line is restricted by genes in the *H-2K* region. Finally, the C57BL/6-derived B6-CSD-35 and B6-CSD-52 hybrids demonstrated NP-specific responses only in *H-2D*- (and *H-2I*-) compatible C57BL/6 and B10.A(2R) recipients.

The present series of *H-2K*-, *H-2I*- and *H-2D*-restricted T-cell hybrids are specific for the same haptenic determinants. Although all the hybrids are NP-specific, the B6-CSI lines do not react with NP-coupled cells and the B6-CSK and B6-CSD lines fail to respond to NP-coupled proteins. This may reflect the interaction requirements for activation of *H-2I*-restricted hybridomas by cells which process and present antigen, while the *H-2D*- or *H-2K*-restricted cells may be activated by a different kind of antigen association[2,12,13]. These data support the hypothesis that activation of the *H-2K*- and *H-2D*-restricted effector CS hybridoma cells is macrophage-independent, as are the interactions of cloned or hybridoma cytolytic T effector cells[14-16].

The series of NP-specific hybridomas isolated and described here provides a very useful tool for analysing antigen receptors on T cells. Further analysis of these hybridoma lines may help to elucidate T-cell function, including analysis of the signals required for cellular activation; comparisons of the products released after antigen activation and determining whether these cells mediate multiple T-cell functions (for example, whether *H-2I*-restricted hybrids may provide helper activity or whether the *H-2K/D*-restricted cells function as killer T cells). Experiments are being done to address these issues.

We thank the Cancer Research Institute and the NIH (AI-16677 and CA-14732) for support, and T. Dinse and N. Axelrod for secretarial assistance.

Received 26 October 1981; accepted 25 March 1982.

1. Sunday, M. E., Weinberger, J. Z., Benacerraf, B. & Dorf, M. E. *J. Immun.* **125**, 1601–1605 (1980).
2. Sunday, M. E. & Dorf, M. E. *J. Immun.* **127**, 766–768 (1981).
3. Okuda, K., Minami, M., Furusawa, S. & Dorf, M. E. *J. exp. Med.* **154**, 1838–1851 (1981).
4. Dennert, G., Weiss, S. & Warner, J. F. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4540–4543 (1981).
5. Lin, Y.-L. & Askonas, B. A. *J. exp. Med.* **154**, 225–234 (1981).
6. Okuda, K., Minami, M., Ju, S.-T. & Dorf, M. E. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4557–4561 (1981).
7. Marcucci, F., Waller, M., Kirchner, H. & Krammer, P. *Nature* **291**, 79–81 (1981).
8. Minami, M., Okuda, K., Furusawa, S., Benacerraf, B. & Dorf, M. E. *J. exp. Med.* **154**, 1390–1402 (1981).
9. Bhattacharya, A., Dorf, M. E. & Springer, T. A. *J. Immun.* **127**, 2488–2495 (1981).
10. Ju, S.-T., Benacerraf, B. & Dorf, M. E. *Proc. natn. Acad. Sci. U.S.A.* **75**, 6192–6196 (1978).
11. Kanno, M., Takei, I., Suzuki, N., Tomioka, H. & Taniguchi, M. *Proc. jap. Soc. Immun.* **10**, 41–42 (1980).
12. Leung, K. N. & Ada, G. L. *Scand. J. Immun.* **12**, 481–487 (1980).
13. Ertl, H. C. J. *Cell. Immun.* **62**, 38–49 (1981).
14. Gillis, S. & Smith, K. A. *Nature* **268**, 154–156 (1977).
15. Reiss, C. S. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 5432–5436 (1980).
16. Kaufmann, Y., Berke, G. & Eshhar, Z. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2502–2506 (1981).

Failure of killed *Listeria monocytogenes* vaccine to produce protective immunity

Carl Heinz Wirsing von Koenig & Horst Finger

Institute of Hygiene and Laboratory Medicine, Municipal Hospital, D-4150 Krefeld, FRG

Herbert Hof

Institute of Hygiene and Microbiology, University of Wuerzburg, D-8700 Wuerzburg, FRG

We report here experiments which, in contrast to recently published data[1,2], suggest that protection against lethal doses of viable *Listeria monocytogenes* cannot be achieved by previous injection with killed *Listeria* vaccine neither by adding macrophage-blocking agents such as dextran sulphate nor by using C3H/HeJ mice having inborn macrophage defects. This discrepancy can be explained by the ineffective killing of bacteria at 56 °C, resulting in injection of small numbers of viable bacteria; such doses are normally subinfective and do not provide immunity. Blocking of the macrophage system by dextran sulphate[3], however, or use of C3H/HeJ mice with inborn macrophage defects[4] renders these doses fully infective, hence providing subsequent protective immunity[5].

L. monocytogenes, serotype 4b, was cultured in tryptose broth at 37 °C for 16 h. The virulence of the strain was maintained by continuous passage in mice of the same strain as used for the experiments. Viable counts of bacteria in broth, dilutions and spleen homogenates were determined using 10-fold dilutions on to agar. Dilutions, where necessary, were made in sterile 0.9% phosphate-buffered saline (PBS) *p*H 7.2.

Using the protocol of van der Meer *et al.*[2], the bacterial suspensions were incubated at 56 °C for 60 min to kill the bacteria. Then the suspensions were diluted to a final concentration of 5.0×10^8 bacteria ml^{-1}. In the first experiment, 0.2 ml of this dilution (1.0×10^8 bacteria) was injected intraperitoneally (i.p.) into female NMRI (Han) mice (Central Institute for Laboratory Animals, Hannover), aged 8–12 weeks. The animals were divided into two groups; one group received an additional i.p. injection of 50 mg dextran sulphate 500 (DS 500) per kg body weight 1 day before infection (Pharmacia). Viable counts of the bacterial suspensions revealed in five independent experiments, between 7.5×10^1 and 8.0×10^2 *L. monocytogenes* per ml, resulting in infective doses between 1.5×10^1 and 1.6×10^2 bacteria per mouse. The time dependence of bacterial killing at 56 °C is shown in Table 1.

The time course of infection of mice was followed by counting bacteria in spleen homogenates at different times after infection. As can be seen from Fig. 1, 1.6×10^2 bacteria per mouse did not result in detectable viable bacteria in the spleens, without further treatment. Challenge of these animals 1 week later with an i.p. injection of $20\times LD_{50}$ (lethal dose for 50% of the population, 4.5×10^5) *L. monocytogenes* led to the death of all animals during the subsequent 5-day observation period. Figure 1 shows that the second group of mice, which were treated with the macrophage blocking agent DS 500, became infected after i.p. injection of only 1.6×10^2 bacteria. The course of infection resembled the normal course of experimental listeriosis[6]. Subsequent challenge of these animals with $20\times LD_{50}$ of viable *L. monocytogenes* revealed protective immunity as only one of eight animals died within the period of observation.

The second experiment used C3H/HeJ and C3HeB/FeJ, male mice (24–28 g; Jackson Laboratories), which have an 'intrinsic adjuvant' (ref. 1). Earlier studies have shown that C3H mice belong to a group of mouse strains that are susceptible to infection with *L. monocytogenes*, a phenomenon thought

Table 1 Killing of *Listeria monocytogenes* at 56 °C

Incubation time (min)	Viable bacteria per ml
0	3.5×10^9
30	2.8×10^6
40	2.8×10^2
50	8.0×10^1
60	1.2×10^2
70	2.0×10^1
80	0
90	4.0×10^1
100	0

L. monocytogenes serotype 4b were grown at 37 °C for 16 h in tryptose broth. The broth was incubated in a 56 °C waterbath then chilled to 4 °C before the viable count was performed.

0028–0836/82/200233—02$01.00

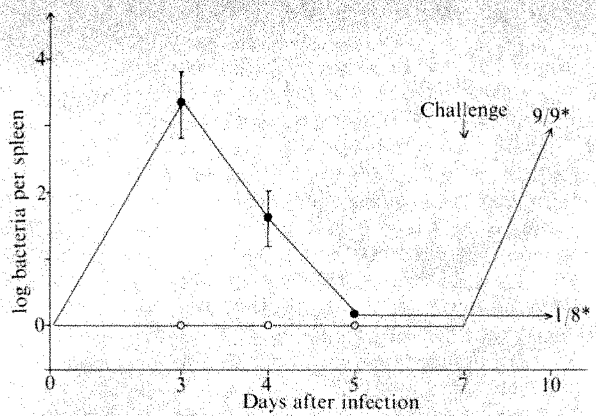


Fig. 1 Course of infection with *Listeria monocytogenes* in DS 500-treated mice. Female NMRI mice were injected i.p. with 1.6×10^2 viable *L. monocytogenes*. ○, Animals given no additional treatment; ●, animals injected i.p. with 50 mg DS 500 per kg body weight on day 1 before infection. Seven days later, all mice were challenged with $20 \times LD_{50}$ (4.5×10^5) viable *L. monocytogenes* i.p. Each point represents the mean ±s.e.m. of five spleens. * Proportion of animals dead at end of observation period.

to be mediated by a single autosomal gene, *Lr*[7]. The HeJ substrain of C3H mice has a macrophage defect[4] which renders it highly susceptible to facultative intracellular parasites such as *Salmonella typhimurium*[8]. Consequently, the i.p. LD_{50} was 4.0×10^3 *L. monocytogenes* compared with 2.3×10^4 for C3HeB/FeJ mice. Both mouse strains were infected with a suspension of bacteria produced as described above. The doses of viable bacteria were fully infective for C3H/HeJ but not for C3HeB/FeJ mice, resulting in induction of protective immunity in the former in the same manner as that shown for DS-500-treated NMRI mice.

We then repeated the experiments using injections of *L. monocytogenes* killed by incubation at 70 °C for 60 min. The killed bacteria were unable to produce immunity or to increase the nonspecific resistance to infection in both DS-500-treated NMRI and in C3H/HeJ mice. We therefore conclude that blocking or impairment of the mononuclear phagocyte system by either DS 500 or *Bordetella pertussis* organisms[9,10], or by inborn defects as in the C3H/HeJ mouse substrain, renders normally subinfective doses of *L. monocytogenes* fully infective, which can be clearly demonstrated by a drastically reduced mean lethal dose. Animals surviving this infection, however, possess a strong protective immunity based on the interaction of T-lymphocytes and macrophages.

These experiments stress the importance of a functionally active macrophage system for resistance against facultative intracellular parasites. Contrary to the findings of van Dijk[1] and van der Meer[2], our experiments suggest that injection with a killed *Listeria* vaccine is unable either to provide protective immunity or to increase nonspecific resistance to infection.

This work was supported by Bundesministerium des Innern.

Received 3 December 1981; accepted 26 March 1982.

1. van Dijk, H., Hofhuis, F. M. A., Berns, E. M. J. J., van der Meer, C. & Willers, J. M. N. *Nature* **286,** 713–714 (1980).
2. van der Meer, C., Hofhuis, F. M. A. & Willers, J. M. N. *Nature* **269,** 594–595 (1977).
3. Hahn, H. & Bierther, M. *Infect. Immunity* **10,** 110–119 (1974).
4. Rosenstreich, D. L. & Vogel, S. N. in *Microbiology 1980,* 11–15 (ASM, Washington, D.C., 1980).
5. Hasenclever, H. F. & Karakawa, W. W. *J. Bact.* **74,** 584–586 (1957).
6. Mackaness, G. B. *J. exp. Med.* **116,** 381–406 (1962).
7. Skamene, E., Konghavn, P. A. L. & Sachs, D. H. *J. infect. Dis.* **139,** 228–231 (1979).
8. O'Brien, A. D. *et al. J. Immun.* **124,** 20–24 (1980).
9. Hof, H., Emmerling, P., Finger, H. & Wirsing, C. H. *Zentbl. Bakt. Hyg. I. Abt.* **A240,** 208–214 (1978).
10. Finger, H., Heymer, B., Wirsing, C. H., Emmerling, P. & Hof. H. *Infect. Immunity* **19,** 950–960 (1978).

Directed effector cells selectively lyse human tumour cells

Charles B. Simone

Clinical Pharmacology Branch, Division of Cancer Treatment, National Cancer Institute, Bethesda, Maryland 20205, USA

Classical antibody-dependent cellular cytotoxicity (ADCC) involves effector cells that mediate damage of antibody-coated targets. Classical ADCC is probably responsible, at least in part, for *in vivo* allograft rejections, resistance to viral infections, parasite destruction and the rejection of tumours[1,2]. Like complement[3,4], human effector cells mediate damage by creating a pore or channel in the target membrane, but the channel created by effector cells is 2.5 times larger (functional diameter ~165Å) than the channel formed by complement[3,5]. Classical ADCC is inhibited by low concentrations of immune complexes which compete with target-bound antibody for effector cell Fc receptors[6]. Because many clinically important diseases, such as certain cancers and systemic infections, have been shown to result in circulating antigen–antibody complexes, it seems likely that these immune complexes might inhibit the ADCC component of host defense mechanism in these diseases. A major advantage to the host might result from the attachment of antibody to effector cells first; either via the Fc receptor[7], which is a weak interaction; by antibody 'associating' with effector cells[8], which is very transient; or by binding antibody to the cellular membrane of effector cells[9]. Here I report the specificity of such antibody-directed ADCC effectors for antigen-bearing targets *in vitro*.

Antibody can be attached to effector cells by incubating them together in the presence of polyethylene glycol (PEG) and spinning them through a mixture of phthalate oils (see Table 1 legend). Using radiolabelled antibody, it can be shown that 150,000 antibody molecules can be attached to each effector cell (neutrophil) by this technique (data not shown). Also, no radiolabelled antibody is detected for up to 24 h in supernatants of directed effectors incubated at 37 °C; or for up to 6 h in the plasma of mice (five animals) which were injected intravenously with directed lymphocytes (data not shown). The effects of

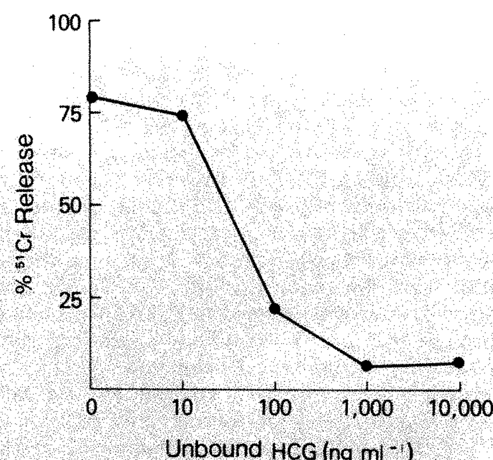


Fig. 1 Effect of unbound HCG on directed effector cell action on ^{51}Cr-labelled JEG-3 cells. 7.5×10^5 ^{51}Cr-labelled -3 cells in RPMI 1640 medium were suspended with 5% fetal calf serum in flat-bottom microtitre wells with indicated amounts of unbound HCG. Anti-HCG-directed effector cells were introduced into the well to a total volume of 200 µl and the plate was gently rotated for 10 min, then incubated for 5 hr at 37 °C. ^{51}Cr release was assayed as before. Controls were effector cells treated with PEG and phthalate oils without antibody, which yielded 7–10% ^{51}Cr release at all points of unbound HCG.

0028–0836/82/200234—03$01.00

Table 1 Action of anti-TNP-directed effector cells on TNP-modified targets

Effector	Anti-TNP ($1\ \mu g\ ml^{-1}$)	% Marker release TNP-ghosts (CF)	TNP-^{51}Cr-RBC	TNP-^{51}Cr-MCF-7
2×10^6 PBLs	0	11	8	10
1×10^6 Neutrophils	0	13	7	7
2×10^6 PBLS*	0	13	9	10
1×10^6 Neutrophils*	0	12	11	11
Directed ADCC				
2×10^6 PBLs	+	94	70	67
$1+10^6$ Neutrophils	+	101	77	76
Classical ADCC				
2×10^6 PBLs	+	96	68	68
1×10^6 Neutrophils	+	93	72	76

Human resealed erythrocyte ghosts containing CF were made as described earlier[3]. Intact RBCs and MCF-7 cells were TNP-modified and ^{51}Cr-labelled. Samples were counted in a γ-counter. Monolayers of each target were made on the bottom surface of flat-bottom microtitre wells as previously described[3]. Freshly separated human granulocytes and lymphocytes were obtained as described previously[3] and used for classical and directed ADCC. They were washed three times and resuspended in balanced salt solution (BSS)–HEPES to a final concentration of 5×10^7 cells per ml. 'Directed' ADCC effector cells were prepared as follows. Affinity-purified rabbit anti-TNP IgG was mixed with effector cells. 12% of 20,000-molecular weight PEG (Fisher Scientific) in BSS–HEPES was immediately added, bringing the total mixture to $1\ mg\ ml^{-1}$ immunoglobulin in 6% PEG. The mixture was incubated for 90 min at 0 °C and then 250 μl aliquots were layered over 800 μl of a phthalate oil mixture (dibutyl phthalate:bis-(2-ethylhexyl)phthalate, both from Eastman Chemicals, Rochester; 1.5:1 v/v) in 1.5 ml Beckman microcentrifuge tubes and sedimented in a Beckman Microfuge B at 9,360 g for 4 min. Cells can be separated from the incubation medium without an aqueous wash because the oil mixture is less dense than the cells and more dense than the medium. The supernatant and entire oil mixture were recovered by aspirating them from the tubes. The cells were washed five times using BSS-HEPES and resuspended to the original concentration of 5×10^7 cells per ml. Control effector cells* were treated with PEG and the phthalate oils as above but without antibody. The microtitre wells were all refilled with BSS-HEPES and the desired reaction components (effector cell, antibody) were added to a final volume of 200 μl. The plates were centrifuged at 250 g for 10 min and incubated at 37 °C for 3 h; then 100 μl aliquots were withdrawn from each well and fluorescence or ^{51}Cr release determined. The total releasable marker was determined by the marker released by 1% Triton X-100. Samples were run in duplicate and agreement between the duplicates was within 15%.

classical ADCC were compared with those of 'directed' ADCC (Table 1) using trinitrophenyl (TNP)-modified targets. Peripheral blood lymphocytes (PBLs) or neutrophils were added to anti-TNP-coated targets (classical ADCC), or effector cells with attached anti-TNP antibody ('directed' ADCC) were added to TNP labelled targets. The three targets used were TNP-resealed human erythrocyte ghosts containing carboxyfluorescein (CF), TNP–^{51}Cr-labelled intact human erythrocytes and TNP–^{51}Cr-labelled MCF-7 cells, a human breast cancer cell line[11]; each was put in separate microtitre wells. Effector cells alone or effector cells treated with PEG and the phthalate oils but without antibody produced minimal background release from each target, showing that the PEG and phthalate oil treatment *per se* does not cause target damage. Directed effector cells can mediate marker release from a membrane model (erythrocyte ghosts), intact erythrocytes and even from human breast cancer cell MCF-7. These results are comparable with those seen when classical ADCC methods are used. Hence the attachment of antibody to effector cells using this technique does not inhibit cell-mediated killing.

The specificity of the directed cytotoxic effector cells was assayed by placing two antigenically distinct targets in the same microtitre well: TNP-modified ghosts containing CF and human choriocarcinoma cell line, JEG-3 (gift of Dr Saul Rosen), labelled with ^{51}Cr. The JEG-3 cell line secretes human chorionic gonadotropin (HCG) at a rate of 7,290 IU per 10^9 cells per day; hormone secretion can, however be blocked by antibody against HCG[12]. If anti-TNP-directed-effector cells are used, very high CF release from the TNP-modified ghost population is observed and there is minimal background ^{51}Cr release from the ^{51}Cr-labelled JEG-3 cells (see Table 2). However, if anti-HCG-directed effector cells (anti-HCG antibody given by Dr Hans Hager, Hoffmann-LaRoche) are used, then ^{51}Cr release is high and CF release is minimal, indicating that the JEG-3 cells are killed and the TNP–ghosts are not damaged. The specificity of directed ADCC is comparable with that of classical ADCC. When soluble immune complexes (HCG–anti-HCG) were added to microtitre wells containing 10^6 JEG-3 cells before 10^6 anti-HCG-directed neutrophils, there was 36% ^{51}Cr net release after 4 h of incubation compared with 40% net release using the same system but without immune complexes. This was in contrast to only 7% ^{51}Cr net release from 10^6 JEG-3 cells which were treated with anti-HCG first, then with immune complexes, then 10^6 effector neutrophils. Hence immune complexes do not significantly inhibit directed ADCC but do inhibit classical ADCC.

JEG-3 cancer cells were chosen as targets because the anti-HCG-directed effector cell–JEG 3 system may be a model clinically relevant to human testicular cancer. It is of interest, therefore, to determine whether unbound HCG (analogous to circulating HCG in a patient with testicular cancer) interferes with the lysis of JEG-3 cancer cells mediated by anti-HCG-directed effector cells. Figure 1 shows that $\geq 1{,}000\ ng\ ml^{-1}$ unbound HCG almost totally prevents anti-HCG-directed effectors from lysing ^{51}Cr-labelled JEG-3 cancer cells, probably by binding to the anti-HCG antibody attached to the effector cells. Whereas 100 $ng\ ml^{-1}$ of unbound HCG allows partial ^{51}Cr release, 0 or 10 $ng\ ml^{-1}$ of unbound HCG does not interfere with killing.

Recently, Miller *et al*[13] treated a T-cell leukaemia patient with systemic anti-Leu 1 antibody and a clinical response, albeit transient was demonstrated. Several difficulties encountered in that patient, including clearance of antibody and dosage, might be avoided in future human *in vivo* therapy if the technique described here was used. Hence the anti-human Leu 1 (Becton Dickinson) reagent was used to see whether it would effect lysis of targets *in vitro* after it had been attached to effector cells in the conditions already described (Table 3). Freshly separated PBLs were ^{51}Cr-labelled and divided into two groups, one of which was TNP-modified. When anti-TNP-directed effectors were put into either group, there was significant ^{51}Cr release from the TNP-modified PBLs and only background ^{51}Cr from the unmodified PBL targets. When anti-human Leu 1-directed

Table 2 Directed compared with classical ADCC using anti-TNP or anti-HCG and their action on targets

Effectors	Anti-TNP: TNP-ghosts %CF release	Anti-TNP: JEG 3 cells %^{51}Cr release	Anti-HCG: TNP-Ghosts %CF release	Anti-HCG: JEG 3 cells %^{51}Cr release
Directed ADCC				
2×10^6 PBLs +antibody	83	8	6	47
1×10^6 Neutrophils +antibody	92	7	7	53
Classical ADCC				
2×10^6 PBLs +antibody	92	7	8	49
1×10^6 Neutrophils +antibody	90	8	8	54
No antibody	TNP–Ghosts	JEG cells		
2×10^6 PBLs	11	7		
1×10^6 Neutrophils	11	7		
2×10^6 PBLs*	13	8		
1×10^6 Neutrophils*	14	8		

4×10^5 TNP-modified erythrocyte ghosts containing CF and 4×10^5 ^{51}Cr-labelled JEG-3 cells were suspended together in flat-bottom microtitre wells in RPMI 1640 medium. Desired components were then added and marker release assayed after 5 h incubation at 37 °C. Directed effector cells were made as before and control effector cells* were treated with PEG and phthalate oils without antibody. Anti-HCG is affinity-purified rabbit IgG.

Table 3 ADCC effector cells directed with anti-TNP or anti-Leu 1 antibodies and their effects on TNP-^{51}Cr-labelled PBL or ^{51}Cr-labelled PBL targets

Effector cell Treatment	% ^{51}Cr release TNP-^{51}Cr-PBLs	^{51}Cr-PBLs
Directed ADCC		
1×10^6 Neutrophils*	15	11
1×10^6 Neutrophils+anti-TNP	74	12
1×10^6 Neutrophils+anti-Leu 1	61	56
Classical ADCC		
1×10^6 Neutrophils*	10	13
1×10^6 Neutrophils+anti-TNP	71	13
1×10^6 Neutrophils+anti-Leu 1	59	62

Freshly separated PBLs were ^{51}Cr-labelled and divided into two groups, one of which was TNP-modified. 7.5×10^5 PBLs of each group were put in separate microtitre wells and suspended in 0.1 M BSS–HEPES and desired components were added to a total volume of 200 µl. Directed effectors were made as before and other control effectors* were treated with PEG and phthalate oils but without antibody. After incubation for 5 h at 37 °C, 100 µl supernatant aliquots were removed and ^{51}Cr content assayed.

effector cells were used, ^{51}Cr was released from each group, but slightly less than the amount seen for the anti-TNP-directed effector cell TNP–PBL reaction. Presumably this is because only the T-cell subpopulation is attacked when anti-human Leu 1 reagent is used, whereas T cells, B cells and null cells are attacked in the TNP–PBL group when anti-TNP is used. Classical ADCC results are comparable.

The technique of using PEG and phthalate oils to attach antibody to freshly separated effector cells results in the specific lysis of cells bearing the target antigen. The lysis is specifically blocked by soluble antigen. Soluble immune complexes do not inhibit directed ADCC. The degree of specific target lysis of human cancer cells and model targets using these directed effector cells is comparable with the lysis produced with classical ADCC techniques. The phthalate oils are completely washed off directed effector cells and these cells cause no adverse reactions *in vivo* (mice and one human). With the advent of monoclonal antibodies, this technique makes it theoretically possible to cause the destruction of targets *in vivo* (tumours, infectious agents) provided the antibody–antigen system is well defined. The absolute amount of non-human antibody given to a patient using this technique (and hence the risk of serum sickness and anaphylactic reactions) should be much less than that amount of antibody passively administered to patients with certain tumours and Gram-negative sepsis[14]. Trials using passive administration of non-human antibody in an attempt to produce human tumour regression have been limited[13,15–17] because recent evidence shows that cell-mediated rather than humoral immunity is important in tumour resistance[18,19] and that passive antibody may enhance tumour growth[20]. I have shown here that human effector cells can be selectively directed to lyse human cancer cells and not other adjacent cells.

I thank Drs A. S. Fauci, D. Longo, C. Myers and F. Abramson for helpful comments.

Received 4 November 1981; accepted 26 March 1982.

1. Cerottini, J. C. & Brunner, K. T. *Adv. Immun.* **18**, 67 (1974).
2. Waksman, B. H. in *Mechanisms of Lymphocyte Activation*, 249 (Wiley, New York, 1974).
3. Simone, C. B. & Henkart, P. A. *J. Immun.* **124**, 954 (1980).
4. Giavedoni, E. B., & Chow, Y. M. & Dalmasso, A. P. *J. Immun.* **122**, 240 (1979).
5. Dourmashkin, R. R., Deteix, P. Simone, C. B. & Henkart, P. A. *Clin. exp. Immun.* **42**, 554 (1980).
6. MacLennan, I. C. M. *Clin. exp. Immun.* **10**, 275 (1972).
7. Segal, D. M. & Hurwitz, E. *J. Immun.* **4**, 1338 (1977).
8. Imir, T., Sakela, E. & Makela, O. *J. Immun.* **117**, 1938 (1976).
9. Jones, J. F. & Segal, D. F. *J. Immun.* **125**, 926 (1980).
10. Greenburg, A. H. & Shen, L. *Nature* **245**, 282 (1973).
11. Soule, H. D., Vasquez, E. R., Long, A., Alberts, S. & Brennan, M. J. *J. natn. Cancer Inst.* **51**, 1409 (1973).
12. Kohler, P. O. *et al. Acta endocr. Suppl.* **150–154**, 137 (1971).
13. Miller, R. A., Maloney, D. G., McKillop, J. & Levy, R. *Blood* **58**, 78 (1981).
14. Ziegler, E. J., McCutchan, J. A. & Braude, A. I. *Trans. Ass. Am. Physns* **91**, 253 (1978).
15. Currie, G. A. *Int. J. Cancer* **26**, 141 (1972).
16. Wright, P. W., Hellström, K. E., Hellström, I. & Bernstein, I. D. *Med. Clins N. Am.* **60**, 607 (1976).
17. Nadler, L. M. *et al. Cancer Res.* **40**, 3147 (1980).
18. Brunner, K. T., Mavel, J., Cerottini, J. C. & Chapuis, B. *Immunology* **14**, 181 (1968).
19. Hellström, I. & Hellström, K. E. *Cancer* **34**, 1461 (1974).
20. Moller, G. *Nature* **204**, 846 (1964).

Kinetics of induction and molecular size of mRNAs encoding human interleukin-2 and γ-interferon

Shimon Efrat, Shlomo Pilo & Raymond Kaempfer

Department of Molecular Virology, The Hebrew University – Hadassah Medical School, 91 010 Jerusalem, Israel

T-cell growth factor (TCGF), or interleukin-2 (IL-2), and immune or γ-interferon (IFN-γ) are lymphokines possessing powerful immunoregulatory properties. IL-2 is essential for the *in vitro* proliferation and maturation of certain classes of T lymphocyte[1,2], particularly cytotoxic[3,4] and helper[5] T cells, and also activates natural killer cells[6]. IFN-γ is considerably more active than IFN-α and -β as activator of natural killer cells and as antiproliferative agent, relative to its antiviral activity[7–11]. As a step towards understanding the regulation of expression of these proteins, we report here the induction of mRNA species encoding IL-2 and IFN-γ in mitogen-stimulated normal human lymphocytes and their expression in *Xenopus laevis* oocytes. On microinjection, distinct species of mRNA sedimenting at 10–10.5S and 13–13.5S direct the synthesis and secretion of active IL-2 and IFN-γ respectively. A second species of IL-2 mRNA is consistently revealed, sedimenting at 13–13.5S and comprising about 20% of the total IL-2 mRNA activity. During induction, an initial concomitant rise in IL-2 and IFN-γ mRNA activities is followed promptly by a concomitant decline in the rate of their accumulation.

We used human tonsils as a convenient source of large amounts of lymphocytes. Cultivation of these cells in the presence of phytohaemagglutinin (PHA) leads to the induction of IL-2 (Fig. 1*a*). IL-2 activity is detectable in the medium by 8 h and accumulates at an increasing rate until about 24 h, when a distinct decline in the rate of accumulation is observed. The kinetics of IFN-γ induction by PHA in the same culture (Fig. 1*b*) are remarkably similar to those of IL-2. Inclusion of the tumour promoter 12-*O*-tetradecanoyl phorbol 13-acetate (TPA) together with PHA enhances both IL-2 and IFN-γ levels in the medium throughout the induction period, but even with TPA present, the rate of accumulation of IL-2 and IFN-γ begins to decline after 20–24 h.

We isolated mRNA from cells that had been cultured for 12 h with PHA and microinjected it into *X. laevis* oocytes. As detailed in Fig. 2 legend, this mRNA directed extensive synthesis of IL-2. About 80% of the expressed IL-2 activity was recovered from the surrounding medium, showing that the mRNA directed both synthesis of active IL-2 and its secretion from oocytes. The mRNA preparation also directed the synthesis of active IFN-γ, the bulk of which was secreted (see Fig. 2 legend).

Table 1 *p*H sensitivity of IFN-γ and IL-2 activities in oocyte medium

	Culture medium	10S mRNA	13S mRNA
IFN-γ(U ml^{-1})			
*p*H 7.2	2,560	<40	220
*p*H 3	<40	<40	<40
IL-2 (c.p.m.)			
*p*H 7.2	35,500	42,500	5,800
*p*H 3	29,200	38,400	7,800

Half the incubation medium of groups of 10 oocytes, injected with 10S or 13S mRNA as described for Fig. 3, was dialysed for 16 h against phosphate-buffered saline at *p*H 3 and then for 2 h against buffer at *p*H 7.2. The other half of the incubation medium was dialysed for 18 h against buffer at *p*H 7.2. Aliquots of 24 h-LCM were treated similarly. Assays were performed as described in Fig. 2 legend.

The kinetics of appearance of mRNA for IL-2 and for IFN-γ in lymphocytes stimulated by PHA, as followed by microinjection into oocytes, are illustrated in Fig. 2. Accumulation of IL-2 mRNA activity begins after 4 h and follows a sigmoid course, with a sharp decline in the rate of accumulation at ~16 h. IFN-γ mRNA activity follows a remarkably similar time course (Fig. 2). Comparison with Fig. 1 shows that the levels of mRNA encoding both IL-2 and IFN-γ reach near maximal values well before these activities appear in the medium in significant amounts.

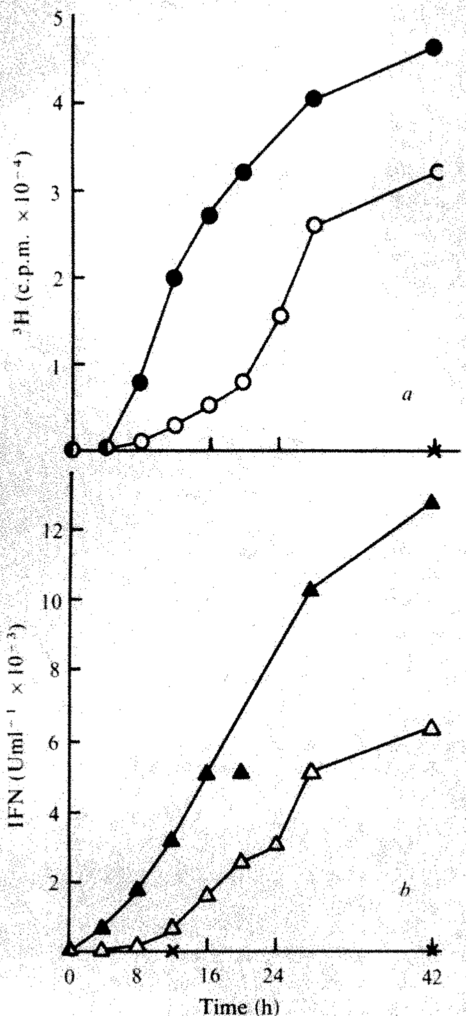


Fig. 1 Kinetics of induction of IL-2 (*a*) and IFN-γ (*b*). Tonsils from otherwise healthy donors were disrupted mechanically and cultured at 4×10^6 cells per ml in RPMI 1640 medium supplemented with 2 mM glutamine, 10 mM minimum essential medium (MEM) non-essential amino acids, 100 mM MEM Na-pyruvate, 10 mM HEPES buffer *p*H 7.2, 5×10^{-5} M 2-mercaptoethanol, 2% fetal calf serum (FCS), antibiotics and 0.4% (v/v) PHA-P (Difco) (○, △). TPA (Sigma) was added to 10 ng ml^{-1} (●, ▲). ×, Mitogen-free controls. At the indicated times, aliquots of the cultures were centrifuged and supernatants were passed through 0.22 μm filters and frozen. *a*, IL-2 activity was assayed by incubating 50 μl of serial dilutions made in culture medium containing 10% FCS with 50 μl of medium containing 2×10^4 assay cells in round-bottom microtitre wells for 48 h at 37 °C. Assay cells were human peripheral blood lymphocytes that had been cultured for at least 1 week in IL-2-containing medium and then incubated without IL-2 for 24 h. Each well received 1 μCi of ^{3}H-thymidine (83 Ci mmol^{-1}) in 10 μl of culture medium at 42 h. The cells were collected on filters and radioactivity in DNA was monitored. Values plotted are for 1:32 dilutions. Values for zero-time cultures, which represent background assay activity of PHA or PHA and TPA, were <500 c.p.m. *b*, IFN-γ was assayed by inhibition of the cytopathic effect of mengovirus in human amnion cells. Confluent cells in microtitre wells were incubated with 50 μl of serial dilutions of the assay fluids in RPMI 1640 medium for 8 h and then infected with 3–5 plaque-forming units (PFU) per cell of mengovirus in 50 μl of medium. After 1 h, the cells were washed and incubated with medium containing 2% FCS for about 24 h before they were stained with Gentian violet. An internal laboratory standard of human IFN-γ and an IFN-α reference standard were included; 1 U of IFN-γ corresponds to about 2.5 IU of IFN-α. The interferon activity was abolished by dialysis against phosphate-buffered saline, *p*H 2.

To estimate the molecular size of IL-2 and IFN-γ mRNAs, mRNA extracted from lymphocytes stimulated for 16 h with PHA was centrifuged through an isokinetic sucrose gradient. In Fig. 3*a*, arrows denote the positions of *Escherichia coli* rRNA and rabbit globin 9S mRNA markers. Microinjection of RNA from individual gradient fractions revealed two peaks of IL-2 mRNA activity, a major one (IL-2A) sedimenting at 10–10.5S and comprising ~80% of the activity, and a minor one sedimenting more rapidly, at 13–13.5S (IL-2B).

Figure 3*b* depicts a sedimentation velocity analysis of mRNA purified from cells stimulated for 16 h with both PHA and TPA and illustrates the synthesis of both IL-2 and IFN-γ directed by individual gradient fractions. Again, IL-2 mRNA activity sedimented in a major peak at 10S and a minor one at 13S, while IFN-γ mRNA activity sediments in a single peak at 13S that directs high levels of IFN-γ synthesis on microinjection into oocytes. The expression of IFN-γ mRNA in these experiments is at least an order of magnitude higher than that described in previous studies[12,13] estimating its size at 15S and 18S respectively. IFN-γ encoded by co-sedimenting mRNA might depress DNA synthesis[11] in the IL-2 assay cells, in which case the amount of the 13S IL-2 mRNA species would be an underestimate. Alternatively, the overlap of the IL-2B and IFN-γ mRNA peaks at 13S suggests that IFN-γ may induce IL-2 in the cells used for the IL-2 assay. To examine these alternatives, we dialysed the medium of oocytes injected with 10S and 13S mRNA at *p*H 3 before testing for IL-2 activity. Such treatment effectively eliminates IFN-γ activity but does not affect the IL-2 activity of the two mRNA peaks (Table 1). We conclude, therefore, that the IL-2B mRNA sedimenting at 13S is expressed independently of IFN-γ.

Inclusion of TPA in the culture medium does not influence the relative abundance of the two IL-2 mRNA peaks, but increases by severalfold the specific activities of the IFN-γ mRNA (see Fig. 2) and IL-2 mRNA species (Fig. 3). The higher levels of induction observed when TPA is present (Fig. 1) are thus reflected by increased mRNA activity.

The mRNA species described here were derived from normal human lymphocytes, avoiding abnormal phenomena that may be associated with transformed cell lines[14]. The ability of IL-2 to stimulate the proliferation of T cells was assayed in cultured human peripheral blood lymphocytes comprising >90% T cells, as judged by rosetting with sheep erythrocytes. These cells are IL-2 dependent and die within 48 h when cultured without this factor. At the time of assay, such cells responded only very weakly to PHA or TPA. Although crude IL-2 preparations may contain B-cell growth factor[15] activity in addition to TCGF, the activity observed in our experiments stimulated extensive DNA synthesis in cells that were essentially pure T lymphocytes, thus defining the activity as IL-2.

The expression of human IL-2 mRNA in *X. laevis* oocytes and the secretion of active IL-2 from these cells made it possible to analyse the levels of IL-2 mRNA activity during the course of induction. Mitogenic stimulation of lymphocytes leads to an initial increase in IL-2 mRNA activity that quickly levels off (Fig. 2). The decline in rate of accumulation of IL-2 mRNA activity agrees well with the decline in rate of IL-2 accumulation observed in the culture medium (Fig. 1). It is thus likely that the capacity of an IL-2 producer cell to form active IL-2 mRNA is limited and that a refractory period may follow, similar to that observed for IFN-β induction[16]. The kinetics of appearance of both IFN-γ (Fig. 1) and IFN-γ mRNA (Fig. 2) are remarkably similar to those for IL-2. The levelling off of mRNA activities observed in Fig. 2 could be caused by the cessation of mRNA synthesis, by the accumulation of mRNA encoding post-transcriptional repressors, or by both. Gullberg *et al.*[17]

have shown that the IL-2-forming capacity of mouse splenocytes declines ~20 h after exposure to mitogen. The present findings support the concept that the induction of IL-2 and IFN-γ in lymphocytes is regulated at the level of transcription.

Finally, our results reveal that IL-2 mRNA is composed of two distinctly sedimenting molecular species (10S and 13S). The mRNA had been treated with dimethyl sulphoxide (DMSO) and heated before centrifugal analysis. This and the observation of a single mRNA peak for IFN-γ, as opposed to two for IL-2, do not support the interpretation that the 13S IL-2 mRNA peak is due to aggregation. This finding probably reflects the existence of more than one gene (or class of genes) encoding IL-2 activity, as has been observed for IFN-β[18,19], although an alternative mRNA processing pathway could also explain the results. The sedimentation coefficient of 10–10.5S for the major IL-2A mRNA species corresponds[20] to a nucleotide length of 610–680 bases. Assuming a minimum of 200 nucleotides for the 5'-leader and 3'-untranslated region, including poly(A), the IL-2A mRNA species can thus encode 135–160 amino acids, corresponding to proteins with molecular weights (MWs) of 14,800–17,600 and more if some side chains are modified. This estimate agrees well with the MW of 15,000 reported for IL-2 (refs 21, 22). The minor IL-2B mRNA species and IFN-γ mRNA correspond to a length of 1,060–1,150 nucleotides and accordingly can encode 31,000–35,000 of unmodified protein. These values are in accord with recent estimates of 20–25,000-MW for IFN-γ[23].

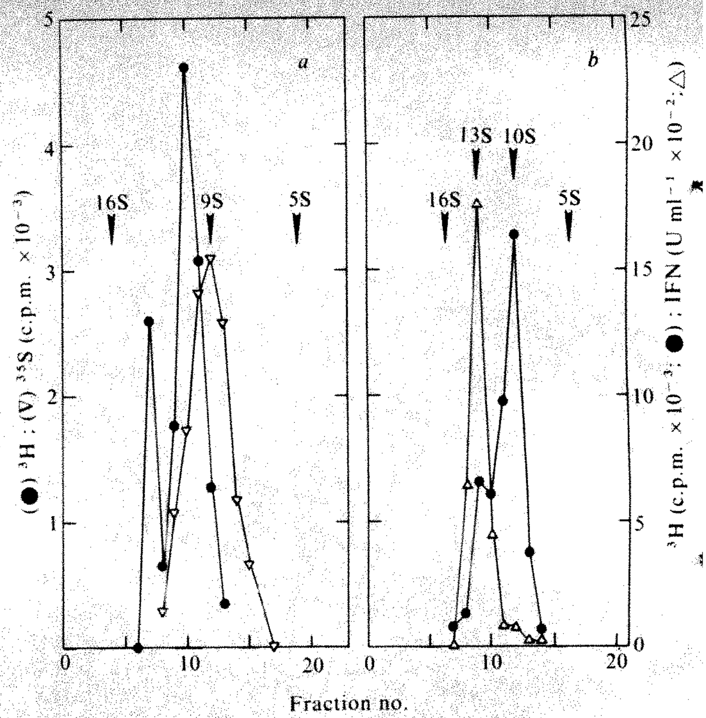


Fig. 3 Sedimentation velocity analysis of IL-2 and IFN-γ mRNAs. Total poly(A)-containing RNA was prepared from lymphocytes cultured for 16 h with PHA (*a*) or PHA and TPA (*b*) as described for Figs 1 and 2. The mRNA (160 μg in *a* and 80 μg in *b*) was heated in water for 90 s at 95 °C, rapidly cooled and centrifuged through 5–17.7% isokinetic sucrose gradients (5 ml) in 10 mM Tris-HCl, *p*H 7.4, 1 mM EDTA and 0.15 M NaCl. RNA from individual fractions was ethanol-precipitated, dissolved in 10 (*a*) or 5 (*b*) μl of water, and microinjected into 10 oocytes as described for Fig. 2. After incubation, the oocyte medium was assayed for IL-2 (●) and IFN-γ (△), as described for Fig. 1. Values for IL-2 are at a 1:4 dilution; at this dilution, standard 24 h-LCM gave 19,000 c.p.m. Assay backgrounds were as for Fig. 2. The concentration of microinjected RNA in the 10S peak fraction of IL-2 was 0.25 (*a*) and 0.21 (*b*) mg ml^{-1}. Parallel gradients contained *Escherichia coli* rRNA A_{260} markers (arrows) or rabbit globin mRNA, monitored for translation activity[27] by incorporation of ^{35}S-methionine (∇).

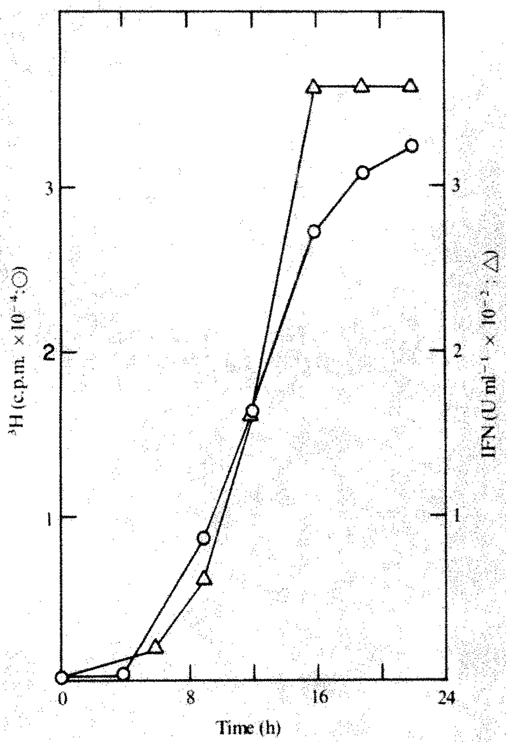


Fig. 2 Kinetics of appearance of IL-2 (○) and IFN-γ (△) mRNA activities in stimulated lymphocytes. Tonsillar lymphocytes were cultured with PHA as described for Fig. 1. At the indicated times, total RNA was extracted by the guanidinium thiocyanate-CsCl method[24]. The RNA was passed twice over oligo(dT)-cellulose, the second time after incubation for 5 min at 55 °C in 90% DMSO[25]. Poly(A)-containing RNA was dissolved in water at a concentration of 2 mg ml^{-1}. Oocytes were each injected with 50 nl of mRNA and incubated for 40 h at 21 °C in groups of 10 with 0.1 ml Barth's medium containing 10^{-4} M phenylmethylsulphonylfluoride[26] (buffer A). Aliquots of incubation medium from each group of oocytes were assayed for IL-2 and IFN-γ activities as described for Fig. 1. Values of IL-2 activity are for 1:4 dilutions. At this dilution, a standard lymphocyte-conditioned medium prepared after 24 h of culturing with PHA (24 h-LCM), gave a count of 30,000 min^{-1}; incubation medium from uninjected oocytes or from oocytes injected with Barth's medium gave <300 c.p.m. background activity. Medium and homogenate (prepared in 0.1 ml of fresh buffer A and cleared by centrifugation) from oocytes each injected with 100 ng of mRNA prepared at 12 h were assayed for IL-2 and yielded 16,400 c.p.m. and 4,400 c.p.m. respectively. When assayed for IFN-γ, these yielded 160 and 25 U ml^{-1} respectively; uninjected oocytes or oocytes injected with Barth's medium gave no detectable background activity. Interferon activity was abolished by dialysis against *p*H 2 buffer.

We thank Dr E. Kedar for helpful discussion and assay cells, Dr H. Soreq for advice on microinjection, and S. Marsh, A. Yeheskel and E. Chriqui for assistance. This work was supported by a grant from the Robert Szold Foundation.

Received 31 December 1981; accepted 24 March 1982.

1. Morgan, D. A., Ruscetti, F. W. & Gallo, R. C. *Science* **193**, 1007–1008 (1976).
2. Gillis, S. & Smith, K. A. *Nature* **268**, 154–155 (1977).
3. Gillis, S., Baker, P. B. Ruscetti, F. W. & Smith, K. A. *J. exp. Med.* **148**, 1093–1098 (1978).
4. Zarling, J. M. & Bach, F. H. *Nature* **280**, 685–687 (1979).
5. Tees, R. & Schreier, M. H. *Nature* **283**, 780–782 (1980).
6. Henney, C. S., Kuribayashi, K., Kern, D. E. & Gillis, S. *Nature* **291**, 335–338 (1981).
7. Claeys, H., van Damme, J., de Ley, M., Vermylen, C. & Billiau, A. *Br. J. Haemat.* **50**, 85–94 (1982).
8. Crane, J. L. Jr, Glasgow, L. A., Kern, E. R. & Younger, J. S. *J. natn. Cancer Inst.* **61**, 871–874 (1978).
9. Sonnenfeld, G., Mandel, A. D. & Merigan, T. C. *Cell. Immun.* **40**, 285–293 (1978).
10. Blalock, J. E., Georgiades, J. A., Langford, M. P. & Johnson, H. M. *Cell. Immun.* **49**, 390–394 (1980).

11. Rubin, B. Y. & Gupta, S. L. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5928–5932 (1980).
12. Taniguchi, T., Pang, R. H. L., Yip, Y. K., Henriksen, D. & Vilcek, J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3469–3472 (1981).
13. Wallace, D. M., Hitchcock, M. J. M., Reber, S. B. & Berger, S. L. *Biochem. biophys. Res. Commun.* **100**, 865–871 (1981).
14. Gillis, S. & Watson, J. *J. exp. Med.* **152**, 1709–1719 (1980).
15. Ford, R. J. *et al. Nature* **294**, 261–263 (1981).
16. Paucker, K. & Boxaca, M. *Bact. Rev.* **31**, 145–184 (1967).
17. Gullberg, M. Ivars, F., Coutinho, A. & Larsson, E. *J. Immun.* **127**, 407–411 (1981).
18. Weissenbach, J. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 7152–7156 (1980).
19. Sehgal, P. B. & Sagar, A. D. *Nature* **288**, 95–97 (1980).
20. Spirin, A. S. *Biokhimiya* **26**, 454–463 (1961).
21. Mier, J. W. & Gallo, R. C. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6134–6138 (1980).
22. Robb, R. J. & Smith, K. A. *Fedn Proc.* **40**, 1163 (1981).
23. Yip, Y. K. *et al. Science* **215**, 411–413 (1982).
24. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. *Biochemistry* **18**, 5294–5299 (1979).
25. Bantle, J. A., Maxwell, I. H. & Hahn, W. E. *Analyt. Biochem.* **72**, 413–427 (1976).
26. Soreq, H. and Miskin R. *FEBS Lett.* **128**, 305–310 (1981).
27. Di Segni, G., Rosen, H. & Kaempfer, R. *Biochemistry* **18**, 2847–2854 (1979).

The *LSP-2* gene and a 5′ flanking sequence are independently expressed in *Drosophila melanogaster*

Barry Yedvobnick* & Michael Levine†

* Department of Biology and
† Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06511, USA

The positions of nucleotide sequences involved in the stage- and/or tissue-specific expression of structural genes in higher organisms are largely unknown. One might predict, however, that closely linked genetic loci which are not coordinately controlled will contain regulatory sequences in close proximity to the respective coding regions. During the third instar of larval development in *Drosophila melanogaster*, the hormone ecdysterone augments the expression of several genes in fat body tissue, one of which is the structural gene coding for larval serum protein-2, *LSP-2*[1–5]. Although *LSP-2* transcripts are detected only in third-instar larval fat bodies, a genomic clone containing the *LSP-2* gene (clone *104*) hybridizes to poly(A)$^+$ RNA prepared from other tissues. We have now analysed clone *104* for other coding regions. Three transcripts, designated *T1*, *T2* and *T3*, hybridize within the 12.5 kilobases (kb) of DNA flanking the 5′ end of the *LSP-2* coding region. The *T1* coding region is separated from the 5′ end of the *LSP-2* coding region by no more than 2.5–3.5 kb. In contrast to the dramatic increase in the level of *LSP-2* transcript induced by ecdysterone in third-instar larval fat bodies, the *T1* transcript remains below the detectable level, indicating that expression of the two coding regions is not coordinately regulated.

A restriction map for clone *104* (Fig. 1 and ref. 3) shows the cleavage sites of *Eco*RI, *Sal*I and *Bam*HI endonucleases and the relevant fragments generated by digesting *104* DNA with either a combination of *Eco*RI and *Sal*I (fragments 1–5) or *Bam*HI (fragments 6–12). The map location of the 5′ terminus for *LSP-2* mRNA was determined by S_1 nuclease digestion[6]. The results of this analysis indicate that the 5′ end of the *LSP-2* coding region is 1.8 kb from the *Sal*I cleavage site in the coding region (Fig. 1). However, we cannot rule out the possibility that the *LSP-2* gene has an intervening sequence which separates a small segment of mRNA coding information from the bulk of the gene. The 1.8 kb value is therefore a minimum estimate.

The initial tests for identifying transcripts that hybridize to *104* DNA involved digestion of the DNA with either a combination of *Eco*RI and *Sal*I or *Bam*HI, separation of the DNA fragments by electrophoresis in agarose gel, transfer of the fragments from the gel to a sheet of nitrocellulose and hybridization with ^{32}P-labelled RNA probes (Fig. 2). Intense hybridization to the fragments containing the *LSP-2* coding region (fragments 2, 3, 9, 12) was detected with the fat body RNA probe but not with the oocyte, gastrula or late embryonic RNAs, in agreement with earlier results indicating that the *LSP-2* gene is expressed specifically in third-instar larval fat bodies[2]. Hybridization was also detected with all four RNA probes to fragments in the region flanking the 5′ end of the *LSP-2* coding region (fragments 1, 4, 7, 8, 11); fragments 7 and 2 hybridized very weakly with fat body and gastrula RNAs respectively; however, the bands are not visible with the autoradiographic exposures shown in Fig. 2. Hybridization to fragment 11 was visible with the fat body RNA probe with shorter exposures. No hybridization was detected with any of the RNA probes to DNA fragments containing sequences flanking the 3′ end of the *LSP-2* coding region.

The transcripts hybridizing with *104* DNA were further characterized by electrophoresis of poly(A)$^+$ RNA from late embryos and total RNA from fat bodies in agarose gel, transfer of the fractionated RNA to nitrocellulose paper and hybridization with ^{32}P-labelled *104* DNA probes (Fig. 3). When intact *104* DNA was used as the probe, a single transcript correspond-

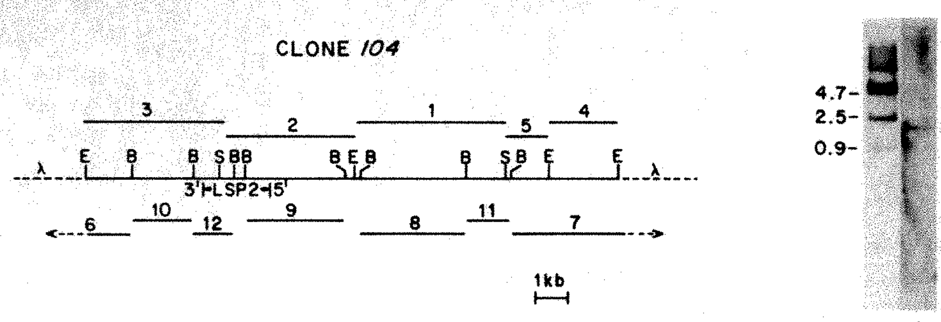


Fig. 1 Restriction map of clone *104*. The clone was isolated from a library of Charon-4 phages containing randomly sheared fragments of *Drosophila* DNA from strain *Canton-S*[12], using as probe poly(A)$^+$ RNA from late third-instar larval fat bodies[1,3]. The restriction map shows the positions in *104* DNA of the endonuclease cleavage sites for *Eco*RI (labelled E), *Sal*I (labelled S) and *Bam*HI (labelled B)[3]. The relevant fragments produced by digestion of *104* DNA with both *Eco*RI and *Sal*I are labelled 1–5, and by digestion with *Bam*HI are labelled 6–12. The location of the 5′ terminus of *LSP-2* mRNA was determined by hybridizing fat body RNA with purified fragment 2 DNA, removing the single-stranded regions by digestion with S_1 nuclease[6] and measuring the size of the remaining RNA–DNA hybrid by electrophoresis in agarose gel. Fragment 2 was purified after electrophoresis in 0.8% agarose gel of an *Eco*RI/*Sal*I digest of *104* DNA, electroelution of the fragment into hydroxyapatite[13] and recovery of the fragment from hydroxyapatite with 1 M sodium phosphate buffer *p*H 7; the purified fragment was dialysed at 4°C against 10 mM Tris/1 mM EDTA *p*H 7.6, precipitated with 95% ethanol and dissolved in 25 μl of a solution containing 0.4 M NaCl/0.04 M PIPES *p*H 6.4/1 mM EDTA/80% formamide[6]. The hybridization reaction was started by dissolving 100 μg total fat body RNA, heating the solution at 67 °C for 10 min and then incubating at 52 °C for 3 h to allow hybrids to form. The solution was then diluted into 10 vols of ice-cold S_1 nuclease reaction buffer and digested with 10 U of S_1 nuclease at 37 °C for 35 min, as described by the manufacturer (BRL). The digestion was stopped and the hybrids precipitated with 95% ethanol in the presence of 10 μg tRNA. The precipitate was electrophoresed in a denaturing agarose gel, transferred to nitrocellulose paper, hybridized to ^{32}P-labelled *104* DNA (specific activity 10^8 c.p.m. μg^{-1}) and autoradiographed, as previously described[3]. The resulting autoradiographs are shown beside the restriction map; lane *a* is a control containing an *Eco*RI/*Sal*I digest of total *104* DNA, lane *b* contains the S_1 nuclease-resistant RNA–DNA hybrid formed with fragment 2. The exposure for *a* was 3 days, for *b* 10 days.

0028–0836/82/200239—03$01.00

ing in size to the *LSP-2* mRNA was detected in fat body RNA but not in embryonic RNA, consistent with earlier results[2]; three other transcripts were detected in the embryonic RNA, of 2.6, 2.2 and 1.6 kb and designated *T1*, *T2* and *T3* respectively. The regions of *104* DNA homologous to each of the three transcripts were identified by using purified *Eco*RI fragments of *104* DNA as individual probes. Fragment 1, which contains the *LSP-2* coding region and 3.5 kb flanking the 5′ end of the coding region, hybridized with the *LSP-2* transcript in fat body RNA and with the *T1* transcript in embryonic RNA; fragment 2 hybridized with transcripts *T1*, *T2* and *T3*; and fragment 3 with only *T3*. The doublet in the lane containing embryo RNA in Fig. 3*b* is probably the result of residual ribosomal RNA in the poly(A)$^+$ RNA preparation. These bands appeared when the blot was hybridized to several probes unrelated to clone *104*. Members of several other laboratories doing similar analyses have found similar, nonspecific bands at the ribosomal RNA position (personal communications).

To investigate whether *T1*, *T2* and *T3* were encoded by separate sequences or shared homology, three additional DNA fragments homologous to *Eco*RI fragment 2 were hybridized to the embryonic RNA blot of Fig. 3 (data not shown, see Fig. 3 legend). We observed that *Eco*RI fragments 2a and 2b were homologous to both *T1* and *T2*, whereas *T3* was homologous

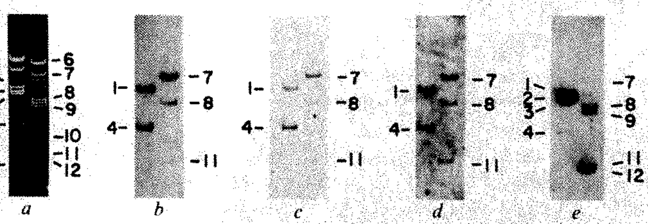


Fig. 2 Regions of *104* DNA that hybridize with RNA from oocytes, embryos and larval fat bodies. Samples containing 1.5 μg of *104* DNA were digested with either a combination of *Eco*RI and *Sal* I or *Bam*HI, separated in 1% agarose gel and transferred to nitrocellulose paper[3]. The paper was prehybridized with 0.75 M sodium chloride/0.075 M sodium citrate/50% formamide/0.2% SDS/0.5 mg ml^{-1} *Escherichia coli* rRNA at 41 °C for 2 h in sealed plastic containers, and the ^{32}P-labelled poly(A)$^+$ RNA probe was added and the reaction continued for 48 h; the RNA was labelled with [γ-^{32}P]ATP and T4 polynucleotide kinase[1], yielding specific activities of 0.5–1.5 × 10^7 c.p.m. μg^{-1}. The paper was washed by shaking for several hours at 41 °C in 0.75 M sodium chloride/0.075 M sodium citrate/50% formamide and then at room temperature in 0.75 M sodium chloride/0.075 M sodium citrate for 30 min. After a 1-min rinse in 0.3 M sodium chloride/0.03 M sodium citrate, the washed paper was wrapped in Saran plastic and autoradiographed for 3–7 days. *a*, Staining pattern with ethidium bromide; *b–e*, autoradiographic patterns btained with poly(A)$^+$ RNA probes from the following sources: *b*, stage 14 oocyte; *c*, gastrula; *d*, late (12 h) embryo; *e*, fat bodies from late third-instar larvae. In all the gels, the left lane contained the *Eco*RI/*Sal*I digest and the right lane contained the *Bam*HI digest of *104* DNA. The positions of the relevant bands are labelled as in Fig. 1.

to *Eco*RI 2c. As probe sequences separated by at least 4 kb are homologous to the 2.6-kb *T1* transcript, the DNA sequences encoding *T1* must be interrupted by at least one intervening sequence. The relationship between the *T1* and *T2* transcripts is uncertain, although the data are consistent with the coding of *T2* by a subset of the *T1* coding region. Additional tests for hybridization of the *LSP-2* and *T1* transcripts with restriction fragment *Eco*RI 1a from *104* DNA place the proximal terminus of the *T1* coding region 2.5–3.5 kb from the 5′ end of the *LSP-2* coding region (data not shown).

Based on the proximity of sequences encoding *LSP-2* and *T1*, we determined whether or not *T1* transcription was augmented during *LSP-2* expression in the fat body. Samples containing poly(A)$^+$ RNA from late embryos and fat bodies were electrophoresed in agarose gel, transferred to nitrocel-

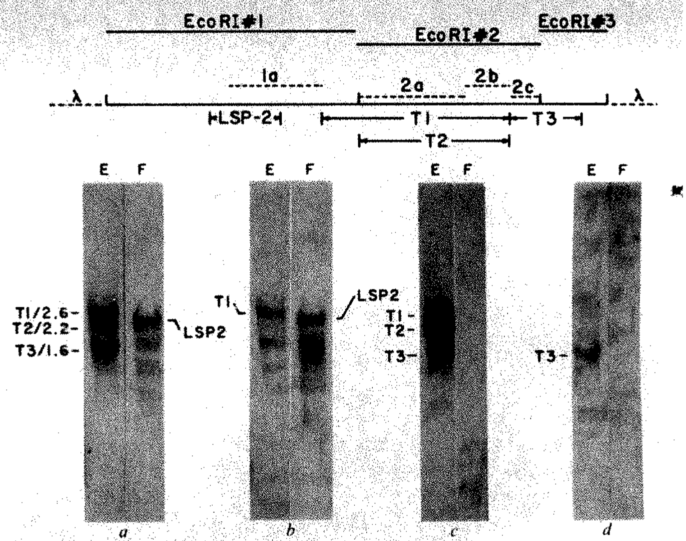


Fig. 3 Identification of transcripts from embryos and larval fat bodies that hybridize with *104* DNA. Samples containing either 23 μg poly(A)$^+$ RNA from 12 h embryos or 10 μg total RNA from late third-instar larval fat bodies were glyoxalated, electrophoresed in an agarose gel and transferred to nitrocellulose paper[3,14]. Hybridizations were done with *Eco*RI fragments of *104* DNA purified by the following procedure. An *Eco*RI digest of *104* DNA was electrophoresed in 1% agarose gel, using low-melting point agarose (BRL), and the DNA bands were detected by staining the gel with ethidium bromide; the bands containing the *Eco*RI fragments of *Drosophila* DNA were cut from the gel and heated at 65 °C for 20 min, and then 100 μl 10 mM Tris-HCl *p*H 7.6/1 mM EDTA containing 15 μg tRNA was added. The solution was brought to 0.2 M sodium acetate and shaken with an equal volume of phenol at 65 °C. The aqueous phase was shaken again with phenol at room temperature; the nucleic acid was precipitated from the aqueous phase in 95% ethanol. Each fragment was labelled by nick-translation, and the labelled probes were hybridized with the RNA transferred to nitrocellulose paper. The hybridizations were done successively using the same paper, with a regeneration step after each hybridization to remove the preceding probe[14]. Probes: *a*, total *104* DNA; *b*, *Eco*RI fragment 1; *c*, *Eco*RI fragment 2; *d*, *Eco*RI fragment 3. E lanes contain late embryo poly(A)$^+$ RNA, F lanes contain total fat body RNA. The map positions of additional probes used in this analysis (1a, 2a, 2b, 2c) are indicated by the dotted lines above the *Drosophila* DNA insert in clone *104*. The results from these hybridizations are explained in the text; the data are not presented. The map positions of the genes encoding *T1*, *T2* and *T3* are indicated by the lines below the *Drosophila* DNA insert.

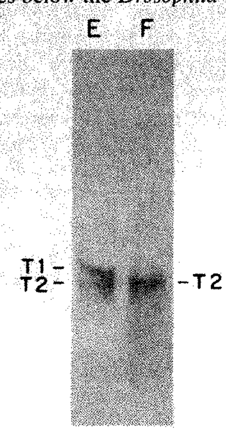


Fig. 4 Test for transcripts *T1* and *T2* in embryonic and fat body RNA. Samples containing 15 μg poly(A)$^+$ RNA from late (12 h) embryos or 12 μg of poly(A)$^+$ RNA from late third-instar larval fat bodies were denatured, electrophoresed in agarose gel and transferred to nitrocellulose paper[3,14]. The *Bam*HI fragment 8 (see Fig. 1) was purified as described for the *Eco*RI/*Sal*I fragment 2 of Fig. 1, ^{32}P-labelled by nick-translation, hybridized to the RNA on nitrocellulose paper, and the paper was then autoradiographed for 3 days. Lane E contains embryonic RNA, lane F contains fat body RNA.

lulose and hybridized to ^{32}P-labelled *Bam*HI fragment 8 (Fig. 4). As predicted, this probe hybridized to the *T1* and *T2* transcripts present in the embryo RNA preparation, whereas only the *T2* transcript was detectable in the fat body sample. When the nitrocellulose filter was regenerated and hybridized to a probe containing the *LSP-2* coding region, the fat body RNA preparation hybridized ~90 times more strongly than the embryo RNA at the position of the *LSP-2* transcript (data not shown). Thus, despite a dramatic increase in the level of *LSP-2* transcript in late third-instar larval fat body tissue, the *T1* transcript does not accumulate.

There is no proof that the sequences encoding the *T1*, *T2* and *T3* transcripts are those flanking the *LSP-2* coding region. *T1*, *T2* and *T3* could be transcribed at other genomic sites, yet possess significant homology with clone *104* sequences. This is unlikely, however, because (see refs 2, 3, 7): (1) when genomic DNA was digested with several restriction endonucleases, electrophoresed in agarose gel and transferred to a nitrocellulose filter, after hybridization to ^{32}P-labelled clone *104* DNA the only bands detected on the filter were those predicted from the restriction endonuclease map of clone *104*. (2) *In situ* hybridizations of tritium-labelled clone *104* DNA to salivary gland chromosomes have revealed only a single site of hybridization, the *LSP-2* locus. Thus, the *T1*, *T2* and *T3* transcripts are probably encoded by the region flanking the *LSP-2* gene. As *T1* transcripts are undetectable in third-instar larval fat body tissue, it seems that the factors which increase the expression of the *LSP-2* gene act locally, rather than regionally. This situation is very similar to the regulation of sequences encoding yeast histones[8], *D. melanogaster* cuticle proteins[9] and the enzyme dopa decarboxylase of *D. melanogaster*[10]. This is particularly interesting since the transcription of *LSP-2* seems to be augmented by the steroid hormone ecdysterone[3,5]. DNA transformation studies have shown that sequences required for hormonal response can be closely linked to the coding regions they regulate[11]. In the case of the *LSP-2* gene, the position of sequences encoding *T1* may delimit the regulatory regions.

B.Y. was supported by a NIH postdoctoral fellowship. M.L. was supported by grants to Alan Garen from the American Cancer Society and the National Institute of General Medical Sciences. This research was performed in the laboratory of Alan Garen.

Received 1 October 1981; accepted 22 March 1982.

1. Lepesant, J. A., Kejzlarova-Lepesant, J. & Garen, A. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5570–5574 (1978).
2. Levine, M., Garen, A., Lepesant, J. A. & Kejzlarova-Lepesant, J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2417–2421 (1981).
3. Lepesant, J. A. *et al. J. molec. appl. Genet.* (in the press).
4. Roberts, D. B., Wolfe, J. & Akam, M. E. *J. Insect Physiol.* **23**, 871–878 (1977).
5. Jowett, T. & Postlethwait, J. H. *Nature* **292**, 633–635 (1981).
6. Berk, A. J. & Sharp, P. A. *Cell* **12**, 721–732 (1977).
7. Levine, M. thesis, Yale Univ. (1981).
8. Hereford, L. M., Osley, M. A., Ludwig, J. R. II & McLaughlin, C. S. *Cell* **24**, 367–370 (1981).
9. Snyder, M., Hirsh, J. & Davidson, N. *Cell* **25**, 165–178 (1981).
10. Hirsh, J. & Davidson, N. *Molec. cell. Biol.* **1**, 475–485 (1981).
11. Hynes, N. E., Kennedy, N., Rahmsdorf, U. & Groner, B. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2038–2042 (1981).
12. Maniatis, T. *et al. Cell* **15**, 687–701 (1978).
13. Tabak, H. F. & Flavell, R. A. *Nucleic Acids Res.* **5**, 2321–2332 (1978).
14. Thomas, P. S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201–5205 (1980).

Male and female mouse DNAs can be discriminated using retroviral probes

Sandra J. Phillips, Edward H. Birkenmeier, Robert Callahan* & Evan M. Eicher

The Jackson Laboratory, Bar Harbor, Maine 04609, USA
* Laboratory of Cellular and Molecular Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205, USA

Only two markers have been identified for the mouse Y chromosome: a gene that determines the male transplantation antigen (H-Y)[1] and sequences homologous to a satellite DNA originally isolated from a heterogametic female snake[2]. To investigate the molecular organization of the murine Y chromosome and to understand its role in sex determination, more Y-specific sequences and genes must be identified. We have now compared restriction digests of DNAs from female and male mice using hybridization probes representing two different classes of murine retroviruses. Some viral DNA-containing fragments were present in male but not female DNAs. The male-specific fragments varied in copy number from 1 to as many as 100, and we estimate that in some *Mus* species, as much as 3% of the Y chromosome is composed of retrovirus-related sequences. To our knowledge this is the first report of (1) an association of retrovirus-related sequences with a mammalian Y-chromosome and (2) many copies of retrovirus-related sequences on the same chromosome.

DNAs were prepared from male and female mice representing six distantly related inbred strains of *Mus musculus* and three other *Mus* species[3]: *Mus poschiavinus*, *Mus spretus* and *Mus caroli*. *M. poschiavinus* and *M. spretus* were included because they interbreed with laboratory strains and *M. caroli* because its karyotype is virtually identical to that of *M. musculus*[4].

*Eco*RI digests of mouse DNAs were blotted onto nitrocellulose[5] and hybridized to the long terminal repeat (LTR) of a *M. musculus* amphotropic retrovirus[6]. The LTR sequence, which shares homology with the AKR ecotropic virus LTR[7], was isolated from unintegrated proviral DNA and subcloned in pBR322. This probe was chosen because it represents terminal viral DNA sequences and detects fusion fragments composed of viral and adjacent cellular sequences. Approximately 50 fragments in *M. musculus* DNA hybridize to the LTR (Fig. 1*a*, lanes 1–12), and they probably represent many of the endogenous type C retroviral genomes known to exist in *M. musculus*[8]. In the four inbred strains BALB/cWt, C57BL/10J, NZB/B1NJ and SM/J, a 5.5 kilobase (kb) *Eco*RI fragment was present in male but not female DNA (Fig. 1*a*, lanes 1–4, 7, 8, 11, 12). The relative intensity of this band suggested that it is present in one or a few copies. This dimorphic fragment was not observed in DNA from AKR/J and DBA/2J strains (Fig. 1*a*, lanes 5, 6, 9, 10).

Restriction fragments containing DNA sequences related to the LTR were weakly detected in the closely related species *M. spretus* (Fig. 1*a*, lanes 13, 14) and *M. poschiavinus* (Fig. 1*a*, lanes 15, 16). No male-specific bands were evident. DNA from the more distantly related species *M. caroli* did not hybridize with this probe, which is consistent with previous results[9,10].

A different hybridization pattern was observed when a sequence representing another class of endogenous xenotropic type C retrovirus was used to analyse these same DNAs (Fig. 1*b*). This class (designated type C-I) has been isolated from all species of *Mus* tested, except *M. musculus*, and can be distinguished from other laboratory strains of retroviruses both immunologically and on the basis of nucleic acid sequence homology[10–12]. A C-I virus (designated M720) was isolated from *M. cookii* tissue culture cells and used to infect FEC cells. A clone of unintegrated circular proviral DNA was prepared. In *M. musculus* (Fig. 1*b*, lanes 1–12) and *M. poschiavinus* (Fig. 1*b*, lanes 15, 16), this clone of the entire *M. cookii* type C-I proviral genome hybridized to at least four fragments (14.5, 11.8, 7.5 and 4.3 kb) in male DNAs that were not present in female DNAs. Because the sum of the sizes of the male-specific bands is larger than a viral genome (8.5 kb), we conclude that these four bands are derived from more than one type C-I endogenous virus genome. The relative intensity of hybridization of these fragments to M720 suggests that each is present in multiple copies. Usually multiple-copy fragments represent

0028-0836/82/200241—03$01.00

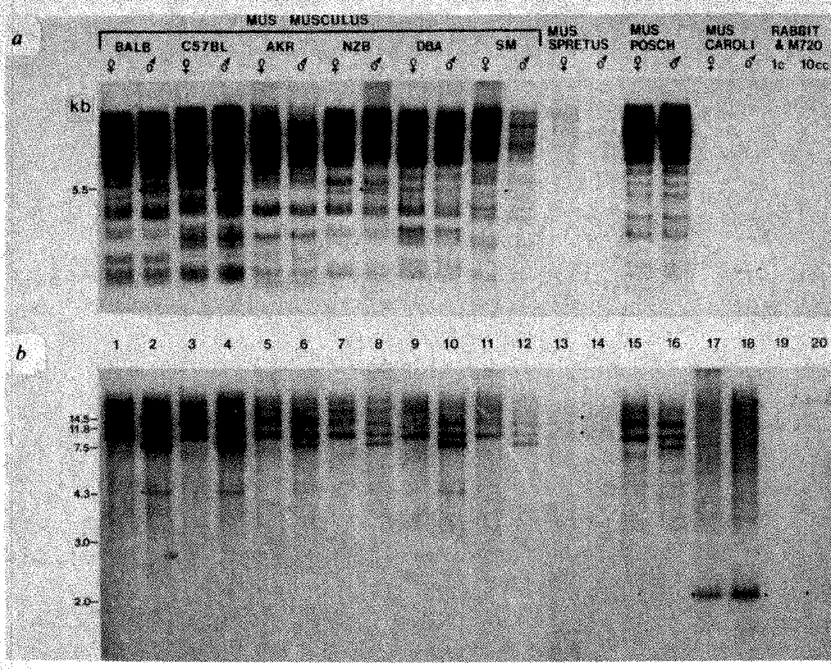


Fig. 1 Detection of sexually dimorphic retrovirus-related *Eco*RI restriction fragments in genomic mouse DNAs. Genomic DNAs were prepared from individual livers of inbred male and female siblings (lanes 1–12) or from closely related animals (lanes 13–18). DNAs (4 μg) were digested with *Eco*RI, separated by electrophoresis in 1% agarose and transferred to nitrocellulose[5]. Probes were prepared by nick-translation to a specific activity of $1\text{–}3\times10^8$ d.p.m. ^{32}P per μg DNA[25]. Filters were incubated at 70 °C in the presence of 6 × SSC for 16 h, and washed extensively at 52 °C in 0.1 × SSC, 0.05% SDS. Hybridizing fragments discussed in the text are denoted by ■ and the sizes of those fregments are given on the left of each panel. In *a*, mouse DNAs were probed with a pBR322 clone of the LTR of an amphotropic retrovirus from *M. musculus*[6]. Resolution was increased by exposing for 10 days to Kodak XAR film without intensifying screens. In *b*, the same DNAs were probed with a pBR322 clone of proviral DNA (M720) from a xenotropic retrovirus from *M. cookii*. Exposure was for 1 day at room temperature. Abbreviated strain and species names include BALB/cWt (BALB), C57BL/10J (C57BL), AKR/J (AKR), NZB/B1NJ (NZB), DBA/2J (DBA), SM/J (SM) and *M. poschiavinus* (MUS POSCH).

internal viral sequences, because flanking cellular restriction sites are presumed to be heterogeneous[13]. If all the *Eco*RI sites generating these multiple-copy fragments are internal viral restriction sites, then we are observing at least one unusually large (>14.5 kb) retrovirus genome. Alternatively, one or more conserved *Eco*RI sites flanking 'normal-sized' viral genomes could generate multiple-copy fragments larger than the viral genome itself.

The number of M720 cirus-related sequences in male DNA from *M. musculus* was estimated by mixing cloned M720 proviral DNA with rabbit DNA at levels representing 1 or 10 copies of provirus DNA per cellular genome (Fig. 1*b*, lanes 19, 20). One copy of M720 proviral DNA per genome is barely detectable when hybridized with itself (lane 19). Ten copies of the viral genome are detectable, but the intensity is approximately one-tenth of that observed with the M720 male-specific fragments shown in Fig. 1*b*, lanes 1–12. From these and other results (data not shown) we conclude that there are ~100 copies of M720 related sequences in males from *M. musculus* inbred strains and *M. poschiavinus*.

Male-specific M720-related fragments were detected in *M. spretus* DNA (Fig. 1*b*, lanes 13, 14); however they hybridize less intensely with the probe than those identified in *M. musculus* male DNA (Fig. 1*b*, lanes 1–12). Whether this reflects a lower copy number or less homology is unresolved. Restricted cellular DNA from *M. caroli* also hybridized with M720 proviral DNA. Although *M. caroli* DNA does not appear to contain male-specific sequences unique to the M720 proviral genome, it does contain a 2.0 kb internal *Eco*RI fragment characteristic of the provirus (Fig. 1*b*, lanes 19, 20). We conclude that the M720-related sequences in the *M. caroli* genome are structurally more closely related to M720 than to related sequences in the other *Mus* species tested. In addition, the overall conserved pattern of the M720-related bands in *M. musculus* and *M. poschiavinus* DNAs is in striking contrast to the strain specificity observed with the LTR probe. These results are consistent with the previously noted conservation of the type C-I endogenous retrovirus-related sequences in cellular DNA from all species of *Mus*[10,12].

The presence of many copies of M720-related sequences on the Y chromosome of some *Mus* species can be explained in several ways. One hypothesis is that one or more independent viral integration events occurred in or near a specific sequence present many times on that chromosome. However, there is no evidence for site-specific integration of retroviruses[14]. Second, multiple copies of a viral genome could have resulted from successive intrachromosomal recombination events. Such a model has been proposed to explain the presence of five copies of the avian leukosis retrovirus genome on a single chicken chromosome[15,16], and would explain the presence of 100 copies of M720-like virus sequences on the mouse Y chromosome. Third, we could be observing the result of amplification of a Y-chromosomal element containing virus-related sequences. The conserved pattern of M720-related dimorphic restriction fragments in *M. musculus* and *M. poschiavinus* suggests that either the integration or amplification event(s) responsible for multiple copies occurred recently in evolutionary terms, or these sequences have been conserved relative to other species. If conservation is responsible, these sequences may have a functional role. The role is not, however, sex determination, because mice that carry a *M. poschiavinus* Y chromosome (as well as the typical M720 male-specific fragments) can develop external female genitalia and ovaries[17]. It may be significant that males of several inbred strains have been shown to express a protein related to a type C virus envelope protein (gp70) in the epididymis and ductus deferens[18,19].

It remains to be demonstrated that the LTR and M720-related patterns of sexual dimorphism represent sequences from the Y chromosome. Although these sequences are present only when a Y chromosome is present, it is possible that the fragment dimorphism is a result of amplification of an autosomal or X-chromosomal sequence. Hybridization *in situ* will provide direct evidence.

We have shown that probes representing dispersed repetitive elements, such as retroviruses, are valuable tools for analysing large genomes. On Southern blots they can effectively reduce the complexity of the murine genome from 1×10^6 *Eco*RI fragments to ~50 fragments. Just as we have used retroviral probes to identify sequences from the Y chromosome, retroviral probes have been useful as molecular markers for identifying regions of the genome where deletions or insertions have occurred[20].

The Y chromosome is unique in the mammalian genome in that it is perpetually monosomic. With the exception of the region that pairs with the X chromosome during meiosis[21], opportunities for recombination with other chromosomes are precluded. Markers such as these virus-related sequences offer a means to study the evolution of this chromosome, which may differ from that of the rest of the genome. Because the Y chromosome represents 2% of the haploid male genome[22,23] (size = 1.5×10^9 base pairs)[24], the M720-related male-specific sequences would represent 3% of the chromosome [(100 copies × 8,500 base pairs/0.02 × 1.5 × 10^9) × 100]. The high concentration of virus-related DNA on the Y chromosome poses

many intriguing questions. Are these viral sequences present as a result of 'poor housekeeping' by the monosomic Y chromosome? Is the Y a 'target' chromosome for retrovirus integration or amplification? Are retroviral sequences involved in the function of this unusual chromosome?

We thank D. W. Bailey, N. A. Jenkins and L. L. Washburn for critically reading the manuscript. The Jackson Laboratory is fully accredited by the American Association for the Accreditation of Laboratory Animal Care. This work was supported by NIH grant GM 20919.

Received 28 January; accepted 15 March 1982.

1. Eichwald, E. J. & Silmser, C. R. *Transplantn Bull.* **2,** 57–58 (1955).
2. Jones, K. W. & Singh, L. *Hum. Genet.* **58,** 46–53 (1981).
3. Marshall, J. T. & Sage, R. D. *Symp. zool. Soc. Lond.* **47,** 15–25 (1981).
4. Miller, O. J. & Miller, D. A. in *Origins of Inbred Mice* (ed. Morse, H. C. III) 591–611 (Academic, New York, 1978).
5. Southern, E. M. *J. molec. Biol.* **98,** 503–517 (1975).
6. Rapp, U. R. & Todaro, G. J. *Proc. natn. Acad. Sci. U.S.A.* **77,** 624–628 (1980).
7. Rands, E., Lowy, D. R., Lander, M. R. & Chattopadhyay, S. K. *Virology* **108,** 445–452 (1981).
8. Jaenisch, R., Fan, H. & Croker, B. *Proc. natn. Acad. Sci. U.S.A.* **72,** 4008–4012 (1975).
9. Chattopadhyay, S. K., Lander, M. R., Rands, E. & Lowy, D. R. *Proc. natn. Acad. Sci. U.S.A.* **77,** 5774–5778 (1980).
10. Benveniste, R. E., Callahan, R., Sherr, C. J., Chapman, V. & Todaro, G. J. *J. Virol.* **21,** 849–862 (1977).
11. Lieber, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **72,** 2315–2319 (1975).
12. Callahan, R. & Todaro, G. J. in *Origins of Inbred Mice* (ed. Morse, H. C. III) 591–611 (Academic, New York, 1978).
13. Steffen, D., Bird, S., Rowe, W. P. & Weinberg, R. *Proc. natn. Acad. Sci. U.S.A.* **76,** 4554–4558 (1979).
14. Weinberg, R. A. *A. Rev. Biochem.* **49,** 197–226 (1980).
15. Tereba, A., Crittenden, L. B. & Astrin, S. M. *J. Virol.* **39,** 282–289 (1981).
16. Tereba, A. *J. Virol.* **40,** 920–926 (1981).
17. Eicher, E. M., Washburn, L. L., Whitney, J. A. III & Morrow, K. E. *Science* (in the press).
18. Del Villano, B. C. & Lerner, R. A. *Nature* **259,** 497–499 (1976).
19. Lerner, R. A., Wilson, C. B., Del Villano, B. C., McConahey, P. J. & Dixon, F. J. *J. exp. Med.* **143,** 151–166 (1976).
20. Jenkins, N. A., Copeland, N. G., Taylor, B. A. & Lee, B. K. *Nature* **239,** 370–374 (1981).
21. Tres, L. L. *J. Cell Sci.* **25,** 1–15 (1977).
22. Committee on Standardized Genetic Nomenclature for Mice *J. Hered.* **63,** 69–71 (1972).
23. Disteche, C. M., Carrano, A. V., Ashworth, L. K., Burkhart-Shultz, K. & Latt, S. A. *Cytogenet. cell. Genet.* **29,** 189–197 (1981).
24. Laird, C. *Chromosoma* **32,** 378–406 (1971).
25. Maniatis, T., Jeffrey, A. & Kleid, D. G. *Proc. natn. Acad. Sci. U.S.A.* **72,** 1184–1188 (1975).

Messenger RNA of infectious pancreatic necrosis virus is polycistronic

P. P. C. Mertens* & P. Dobos

Department of Microbiology, University of Guelph, Guelph, Ontario, Canada N1G 2W1

Infectious pancreatic necrosis virus (IPNV) is an economically important fish pathogen. The virus particles contain two large segments of double-stranded (ds)RNA[1–4], which encode four different intracellular primary gene products[5,6]. The smaller genome segment ('B') encodes the largest of these four proteins and the larger genome segment ('A') encodes the other three by a mechanism which does not seem to involve post-translational cleavage of a large precursor polypeptide *in vivo*[7]. Previous results[8,9] have indicated that only two species of mRNA are produced, one from each genome segment[8], and that one of these viral mRNAs (that produced from segment B; ref. 7) is monocistronic. The other mRNA, which is transcribed from segment A (ref. 7), encodes three proteins, which may be produced by a mechanism involving multiple initiation and termination sites for translation, rather than post-translational cleavage[5–7]. To distinguish between these two possible mechanisms, we have studied the translation *in vitro* of the denatured IPNV genomic dsRNA segments. We report here that the mRNA transcribed from genome segment A is polycistronic.

IPNV particles are icosahedral with a diameter of 60 nm. The dsRNA genome is bi-segmented (segments A and B of molecular weight (M_r) 2.5×10^6 and 2.3×10^6, respectively[1–4]; see Fig. 1*A*).

* Present address: The Animal Virus Research Institute, Pirbright, Woking, Surrey GU24 0NF, UK.

Three size classes of IPNV-specific polypeptides are synthesized in the same relative proportions throughout the infectious cycle, starting 3 h after infection[5]. These primary gene products are designated ICPs (infected cell proteins) and have molecular weights of 105,000 (105K), 62K, 31K and 29K (Fig. 1*C*[5,7]). Of these four proteins, three subsequently undergo post-translational cleavage: (1) the 62K protein is cleaved via a 60K intermediate, to generate the major capsid protein VP54 (Fig. 1*D*); (2) part of the 31K internal virion protein is cleaved during maturation to generate a 29K polypeptide, and both

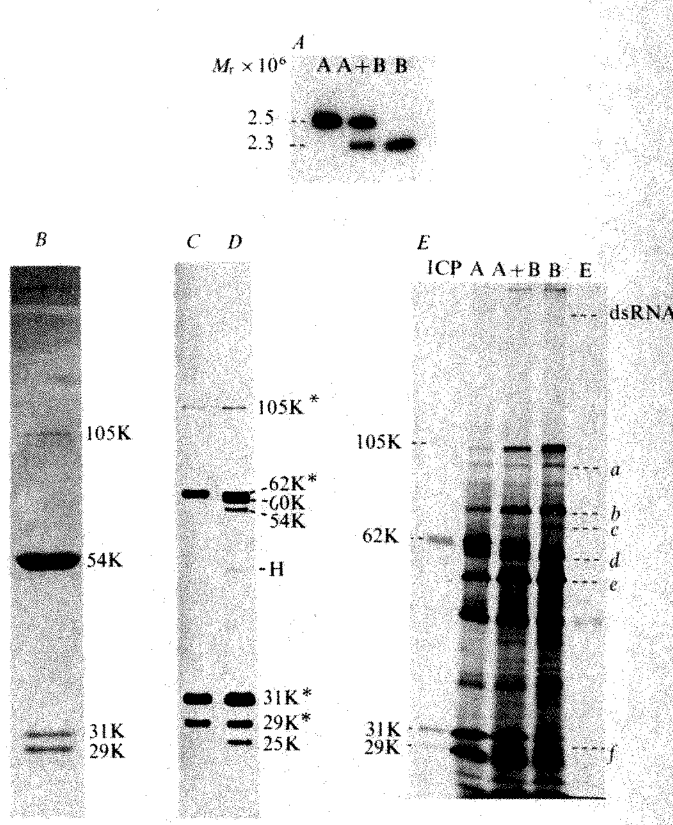


Fig. 1 Polyacrylamide gel electrophoresis of IPNV dsRNA, purified virus proteins, virus-specific proteins synthesized in infected cells, and protein products derived by translation *in vitro* of denatured IPNV dsRNA. Purified virus was digested with proteinase K and the genome segments were separated by electrophoresis in 10% polyacrylamide slab gels[10,17]. The bands were detected by autoradiography (a small amount of ^{32}P-labelled dsRNA was added to the original sample), sliced out and electroeluted from the gel slices using methods published elsewhere[10]. *A*, a small amount of each fractionated genome segment was subjected to analytical gel electrophoresis followed by autoradiography to check its purity and homogeneity (note that due to trailing, segment A was contaminated with segment B). The molecular weights were computed from their electrophoretic mobilities relative to reovirus (Dearing 3) dsRNA run in parallel wells. *B*, proteins of purified IPNV analysed in 10% polyacrylamide slab gel followed by staining with Coomassie blue using procedures published elsewhere[17,18]. *C*, *D*, autoradiograms of labelled virus-specific polypeptides analysed in 11% polyacrylamide slab gels. Infected CHSE cell monolayers were UV-irradiated 3 h after infection to reduce the cellular background and the cultures were labelled with ^{35}S-methionine 6 h after infection at 20 °C for 1 h (*C*) or 4 h (*D*) as described previously[5,6]. The molecular weights of virus-specific proteins were computed from their electrophoretic mobilities relative to unlabelled standards run in parallel. H indicates residual host protein; the four primary gene products are marked with asterisks. *E*, cell-free protein synthesis was carried out in nuclease-treated rabbit reticulocyte lysates (NEN), primed with denatured dsRNA segments, in the presence of ^{35}S-methionine. 2 μg of dsRNA were precipitated in ethanol, washed with ether, dried and denatured in 2 μl of dimethyl sulphoxide at 55 °C for 15 min, then 100 μl of the complete reaction mixture was added and incubation was continued at 37 °C for 1 h. The products of translation were analysed using 11% polyacrylamide slab gels followed by autoradiography. ICP, labelled infected-cell proteins, synthesized in UV-irradiated CHSE cell cultures from 6 to 8 h after infection. A, *in vitro* translation products from dsRNA segment A; A+B, *in vitro* translation products from segments A and B; B, *in vitro* translation products from segment B; E, endogenous products of translation (no RNA added). The approximate molecular weights of the major products are indicated. *a–f*, The major products from incomplete translation of segment B. (Note the location of ^{32}P-labelled dsRNA near the top of the gel.)

0028–0836/82/200243—04$01.00

precursor (VP31) and product (VP29) are found in the virion[6] (Fig. 1*C*); (3) the only non-structural protein, 29K, is slowly degraded into small polypeptides via a 25K intermediate[6] (Fig. 1*D*). The 105K protein is synthesized in small quantities and is incorporated into virions (VP105) without apparent changes in molecular weight (Fig. 1*B–D*). Peptide mapping has shown that the four primary gene products have different amino acid sequences[6].

Using hybrid viruses of two parental serotypes, it has been shown that genome segment B encodes ICP 105K and A encodes the other three primary gene products[7]. The question arises as to how three unrelated proteins are produced from a single large genome segment. We considered the following possible mechanisms: (1) post-translational cleavage of a large, precursor polypeptide; (2) transcription of subgenomic mRNAs from genome segment A; (3) transcription of genome-length mRNA containing multiple initiation and termination sites for protein synthesis.

Several studies have provided evidence against mechanisms 1 and 2[5,6,8,9]. For example, we have been unable to detect polyprotein precursor(s) in IPNV-infected cells by pulse-chase experiments or by using amino acid analogues, protease inhibitors, $ZnCl_2$ or supra-optimal temperatures[5] (experiments with pactamycin also gave negative results; P.D., unpublished results). In addition, the relative molar ratios of the primary gene products indicated that the frequency of their synthesis was inversely related to their molecular weights, suggesting that they are the products of different cistrons[5]. Furthermore, during intracellular replication of IPNV, only two species of genome-length 24S viral mRNAs were synthesized, which hybridized to the two genome segments[8]. The virion-associated polymerase was shown to synthesize *in vitro* two species of single-stranded (ss)RNA which were the same and which co-migrated during electrophoresis with the species of viral message synthesized *in vivo*[9]. Thus no discrete subgenomic mRNAs were detected either *in vivo*[8] or *in vitro*[9].

To investigate the third possible mechanism (that is, multiple initiation and termination sites on a single mRNA species), we used an *in vitro* protein synthesizing system. Denatured virion dsRNA was used to provide translation templates[10] in rabbit

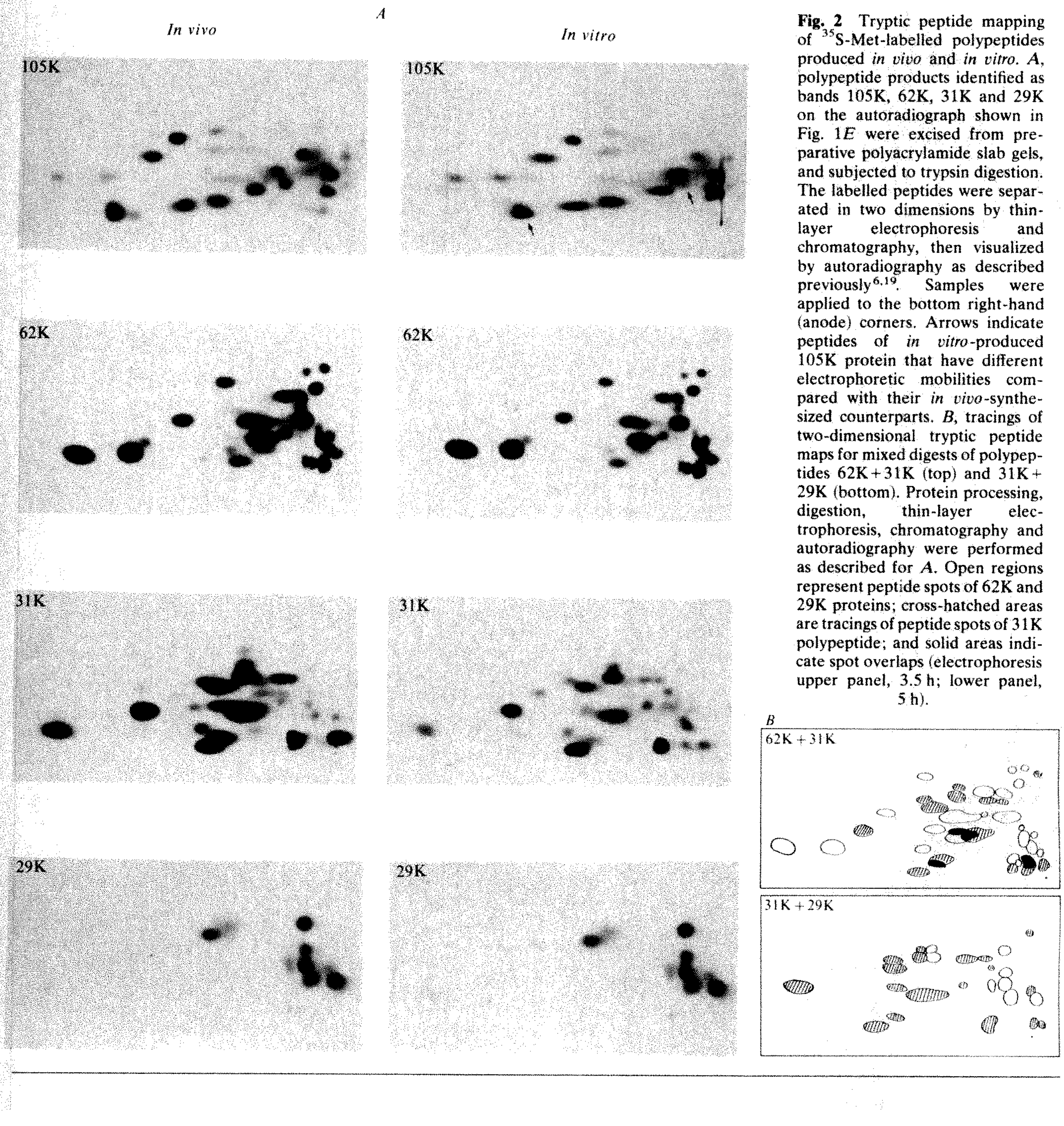


Fig. 2 Tryptic peptide mapping of ^{35}S-Met-labelled polypeptides produced *in vivo* and *in vitro*. *A*, polypeptide products identified as bands 105K, 62K, 31K and 29K on the autoradiograph shown in Fig. 1*E* were excised from preparative polyacrylamide slab gels, and subjected to trypsin digestion. The labelled peptides were separated in two dimensions by thin-layer electrophoresis and chromatography, then visualized by autoradiography as described previously[6,19]. Samples were applied to the bottom right-hand (anode) corners. Arrows indicate peptides of *in vitro*-produced 105K protein that have different electrophoretic mobilities compared with their *in vivo*-synthesized counterparts. *B*, tracings of two-dimensional tryptic peptide maps for mixed digests of polypeptides 62K+31K (top) and 31K+29K (bottom). Protein processing, digestion, thin-layer electrophoresis, chromatography and autoradiography were performed as described for *A*. Open regions represent peptide spots of 62K and 29K proteins; cross-hatched areas are tracings of peptide spots of 31K polypeptide; and solid areas indicate spot overlaps (electrophoresis upper panel, 3.5 h; lower panel, 5 h).

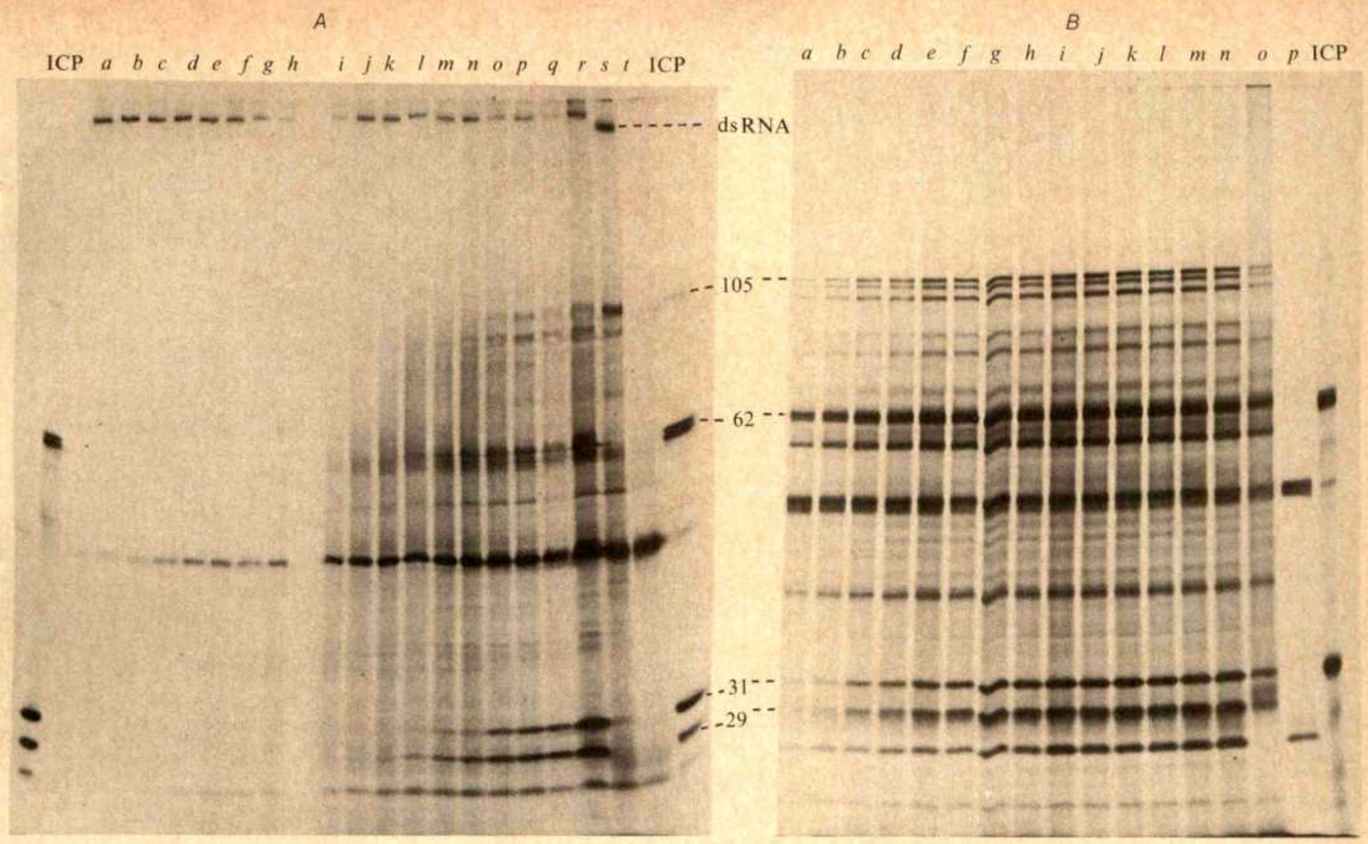


Fig. 3 Analysis of proteins synthesized *in vitro* by translation of denatured IPNV dsRNA in the presence of ^{35}S-methionine. *A*, time course of protein synthesis; cell-free protein synthesis was initiated as described for Fig. 1*E*. At intervals, 2-μl samples were removed from the reaction mixture and denatured by boiling in SDS-containing electrophoresis sample buffer. The labelled products were analysed in 11% polyacrylamide slab gels as described for Fig. 1*E*. *In vitro* translation primed with dsRNA segment A was stopped after the following incubation periods (min): *a*, 1; *b*, 2; *c*, 3; *d*, 4; *e*, 5; *f*, 6; *g*, 7; *h*, 8; *i*, 10; *j*, 12; *k*, 14; *l*, 16; *m*, 18; *n*, 20; *o*, 25; *p*, 30; *q*, 45; and *r*, 60. *s*, *In vitro* translation products from genome segment B, 60 min. *t*, Endogenous product (no RNA added), 60 min. ICP, labelled infected cell proteins as in Fig. 1*E*. (Note the position of ^{32}P-labelled dsRNA segments near the top of the gel.) The approximate molecular weights of the primary gene products are indicated. *B*, pulse-chase experiment; cell-free protein synthesis was initiated as described for Fig. 1*E* using both segments of dsRNA. At various intervals, 5 μl were removed from the reaction mixture, a 1,000-fold excess of unlabelled methionine was added and incubation continued for a total of 60 min. The labelled products were analysed by electrophoresis as described for Fig. 1*E*. Excess unlabelled methionine was added to samples of the reaction mixture at the following times (min): *a*, 2; *b*, 4; *c*, 6; *d*, 8; *e*, 10; *f*, 12; *g*, 14; *h*, 16; *i*, 18; *j*, 20; *k*, 30; *l*, 40; *m*, 50; *n*, 60; and *o*, 60, followed by incubation at 37 °C for 16 h. *p*, No RNA was added (endogenous, 60 min). ICP, labelled infected cell proteins as in Fig. 1*E*.

reticulocyte lysate. The two dsRNA genome segments were separated by preparative polyacrylamide gel electrophoresis (Fig. 1*A*) and recovered by electroelution[10]. Due to trailing, segment A was contaminated with B even after two cycles of electrophoresis, whereas segment B was homogeneous. The results in Fig. 1*E* show that proteins having similar electrophoretic mobilities to the four primary gene products produced in infected cells were also synthesized *in vitro*. Translation of genome segment B produced polypeptide 105K and several smaller polypeptides (labelled *a–f*) which were the result of premature termination of translation, as determined by peptide map comparisons (data not shown). The 105K protein produced *in vitro* exhibited slightly faster electrophoretic mobility than the same protein made in infected cells. This may be due to the addition *in vivo* of prosthetic groups, for example by glycosylation, methylation or phosphorylation, which may have reduced the electrophoretic mobility of ICP 105K. This electrophoretic difference was also reflected in minor variations in the otherwise identical peptide patterns of the *in vivo* and *in vitro* synthesized polypeptide 105K (Fig. 2*A*, arrows). This polypeptide is a minor product *in vivo*[5,6] (Fig. 1*C*), but was produced in large quantities *in vitro* suggesting that not all of the intracellular translational control mechanisms are operative.

Translation of genome segment A resulted in three additional polypeptides having identical electrophoretic mobilities to ICP 62K, 31K and 29K (Fig. 1*E*). Due to contamination of the preparation with the very efficient template B RNA, polypeptide 105K and its incomplete translation products were also found, although in reduced amounts. The tryptic peptide maps of the *in vitro* synthesized proteins 62K, 31K and 29K are identical to those of the equivalent proteins produced in IPNV-infected cells (Fig. 2*A*).

Although rapid protein processing could not be detected *in vivo*[5], the high degree of radiolabelling, and the synchronous initiation of translation achieved *in vitro* prompted us to re-examine the possibility of post-translational cleavage. Analysis of the time course of protein synthesis where *in vitro* translation primed with dsRNA segment A was terminated at intervals between 0 and 60 min of incubation, showed that the order of appearance of polypeptides was 29K, 31K, 62K (Fig. 3*A*). The time for completion of polypeptide chains was proportional to their length, that is, smaller proteins were completed sooner (10 min) than longer ones (16 min). A large molecular weight precursor polypeptide could not be detected. Such a pattern is compatible with three independent initiation sites; it could also arise, however, if the order of polypeptides in a putative precursor were NH_2-29K-31K-62K-COOH and if rapid proteolytic cleavage occurred as soon as the ribosome had passed the 3′ end of a given cistron.

To distinguish between these two alternative mechanisms, we performed the following *in vitro* pulse-chase experiment. Translation was initiated in the presence of ^{35}S-methionine, then at 2-min intervals a 1,000-fold excess of unlabelled methionine was added and incubation continued for 1 h to complete chain elongation. The data in Fig. 3*B* show that all three proteins became labelled even during the shortest pulse

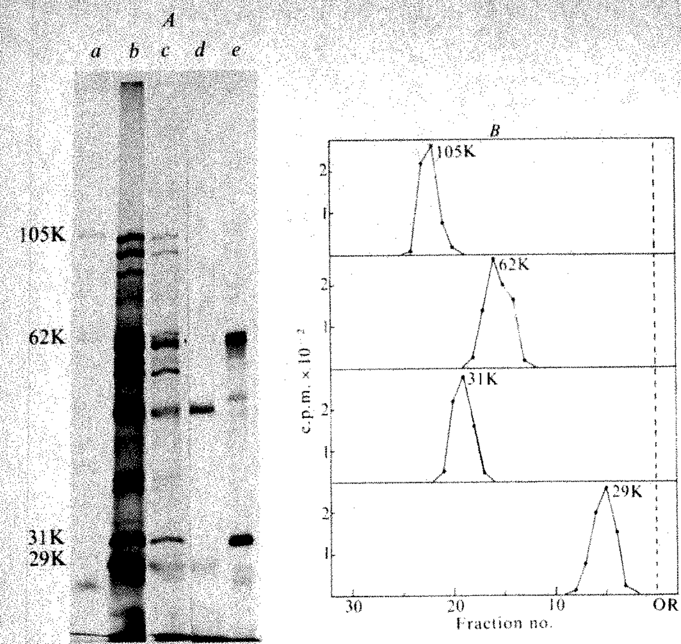


Fig. 4 Analysis of proteins synthesized *in vitro* by translation of denatured IPNV dsRNA in the presence of ^{35}S-fMet-tRNA$_f^{Met}$. *A*, cell-free protein synthesis was carried out using both segments of denatured dsRNA in the presence of ^{35}S-f-Met-tRNA$_f^{Met}$ (as described for Fig. 1*E*). The labelled products were analysed in 11% polyacrylamide slab gels and visualized by fluorography. *a*, ^{35}S-Met-labelled *in vitro* translation product; *b*, ^{35}S-f-Met-labelled *in vitro* translation product; *c*, same as *b* but exposed for a shorter period; *d*, no added RNA (endogenous product); *e*, ICP as in Fig. 1*E*. The molecular weights of the primary gene products are indicated along side. *B*, Thin-layer electrophoresis of the ^{35}S-f-Met-labelled tryptic peptides of polypeptides 105K, 62K, 31K and 29K. Processing and tryptic digestion of fMet-labelled polypeptides were performed as described in Fig. 2 legend. The digests were applied near the anodic end of a thin-layer silica gel plate and analysed in parallel by electrophoresis in pyridine/acetic acid/water (1:10:100 v/v/v) *p*H 3.5 at 450 V for 6.5 h. The dried gel was sliced into 5-mm fractions and each fraction was placed in a toluene-based scintillation cocktail and counted in a liquid scintillation counter. The radioactivity plotted represents net counts greater than twice the background. The broken line marks the origin (OR), where samples were applied. Electrophoresis was from right to left, with the anode on the right.

(2 min). These results are consistent with simultaneous initiation of translation at three independent sites (one for each protein). If initiation had occurred at a single site, followed by cleavage of the nascent chain, then the proteins should have been labelled sequentially from the N-terminus of the putative precursor molecule, with increasing pulse time.

It may be argued that the data are also compatible with the three proteins having common N-termini (overlapping genes), in which case the gene encoding 29K would be situated within the gene for polypeptide 31K, which in turn would be located within the gene for 62K protein. If this were the case, the peptide maps of 29K and 31K proteins would be virtually identical and would show a 50% overlap with that of 62K protein; however, this is not the case as shown in Fig. 2*B* (see also Fig. 2*A*).

When the four proteins were synthesized *in vitro* in the presence of ^{35}S-fMet-tRNA$_f^{Met}$, the N termini of all four polypeptides became labelled with ^{35}S-fMet, indicating independent initiation of translation (Fig. 4*A*). Furthermore, the labelled (N-terminal) tryptic peptides of these proteins exhibited different electrophoretic mobilities when analysed by thin-layer gel electrophoresis (Fig. 4*B*). These results do not, however, exclude the possibility of the three genes initiating and overlapping in the different reading frames on the mRNA.

That intact renatured ^{32}P-labelled dsRNA segments were detected after translation *in vitro* (Figs 1*E*, 3*A*), indicates that there was neither detectable breakdown nor processing of the input denatured dsRNA into smaller monocistronic RNA species. We therefore conclude that the plus strand of genome segment A (that is, the mRNA transcribed from this genome segment) is polycistronic.

Alternative explanations of these results are possible, although in our view highly unlikely. For example it may be argued that (1) IPNV contains more than two qualitatively different RNA segments; however we have shown recently by genetic experiments that the virus contains only two RNA species[7]. (2) One or two splices in the mRNA could result in a change of reading frame without appreciably altering the molecular weight of the RNA, but there is no precedent for this in the case of dsRNA-containing viruses. Although influenza virus Ns_2 mRNA is synthesized from NS_1 mRNA by processing[11], the sizes of the two NS mRNAs are substantially different. It also seems unlikely that the IPNV genomic plus strand would be spliced correctly in reticulocyte lysates *in vitro*. (3) It may also be argued that the *in vivo* mRNA of IPNV is not completely identical to the plus strand of the virus genome and that the two RNAs may behave differently during translation. We believe that only complete sequence analysis of both RNAs can conclusively answer this question, however, partial sequence studies of reovirus indicate that, at least in this case, the viral mRNA and genomic plus strands are identical[12].

Eukaryotic mRNAs usually cannot display more than one initiation site for protein synthesis at a time, even if a second site is present on the mRNA and can be rendered functional by a change in configuration, as in the case of togaviruses[13]. The mRNA of Kunjin virus (a member of the flaviviruses) may contain multiple initiation sites for translation; however, this has not been demonstrated by *in vitro* translation using ^{35}S-f Met labelling, nor has it been confirmed by other investigators[14]. Recently, indirect evidence has indicated that the late 16S mRNA of simian virus 40 is polycistronic, encoding the agnoprotein and VPl (ref. 15).

Although the results given here indicate that translation of the three proteins encoded by genome segment A of IPNV is initiated independently, we have no evidence concerning the relative position of the different cistrons on the mRNA. It is possible that the three coding regions are located in different reading frames and therefore overlap to some extent. Work is in progress to determine the order of cistrons on the A segment of the IPNV genome. We have also begun *in vitro* translation experiments to elucidate the mechanism of gene expression of two other bi-segmented dsRNA-containing viruses, that is, *Drosophila* X virus and infectious bursal virus of chickens[16].

This work was supported by the Natural Sciences and Engineering Research Council of Canada. We thank Drs F. Brown, C. Payne, D. Rowlands, D. Sangar and R. Macdonald and Mr B. Underwood for helpful criticism during the preparation of this manuscript.

Received 17 November 1981; accepted 22 March 1982.

1. Cohen, J., Poinsartd, A. & Scherrer, T. *J. gen. Virol.* **21**, 485–498 (1973).
2. Dobos, P. *Nucleic Acids Res.* **3**, 1903–1924 (1976).
3. Macdonald, R. D. & Yamamoto, T. *J. gen. Virol.* **34**, 236–247 (1977).
4. Dobos, P., Hallett, R., Kells, D. T., Sorensen, O & Rowe, D. *J. Virol.* **22**, 150–159 (1977).
5. Dobos, P. *J. Virol.* **21**, 242–258 (1977).
6. Dobos, P. & Rowe, D. *J. Virol.* **24**, 805–820 (1977).
7. Macdonald, R. D. & Dobos, P. *Virology* **114**, 414–422 (1981).
8. Somogyi, P. & Dobos, P. *J. Virol.* **33**, 129–139 (1980).
9. Mertens, P. P. C., Jamieson, P. J. & Dobos, P. *J. gen. Virol.* **57**, 47–56 (1982).
10. McCrae, M. A. & Joklik, W. K. *Virology* **89**, 578–593 (1978).
11. Inglis, C. S., Gething, M. J. & Brown, C. M. *Nucleic Acids Res.* **8**, 3575–3589 (1980).
12. McCrae, M. A. *J. gen. Virol.* **55**, 393–403 (1981).
13. Kennedy, S. I. T. *J. molec. Biol.* **108**, 491–551 (1976).
14. Westaway, E. G. *Virology* **80**, 320–335 (1977).
15. Gilbert, J., Normura, S., Anderson, C. W. & Khoury, G. *Nature* **291**, 346–349 (1981).
16. Dobos, P. *et al. J. Virol.* **32**, 593–605 (1979).
17. Laemmli, U. K. *Nature* **227**, 680–685 (1970).
18. Carstens, E. B. & Weber, J. *J. gen. Virol.* **37**, 453–474 (1977).
19. Elder, J. A., Pickett, R. A. II, Hampton, J. & Lerner, R. A. *J. Biol. Chem.* **252**, 6510–6515 (1977).

BOOK REVIEWS

The games mathematicians play

C.W. Kilmister

Winning Ways, for Your Mathematical Plays. By Elwyn R. Berlekamp, John H. Conway and Richard K. Guy. Two vols, pp.850. Vol. 1 *Games in General*, hbk ISBN 0-12-091150-7, pbk ISBN 0-12-091101-9; Vol. 2 *Games in Particular*, hbk ISBN 0-12-091152-3, pbk ISBN 0-12-091102-7. (Academic: 1982.) Each volume hbk £31.40, $64.50; pbk £10.80, $22.50.

THE business of games is too serious to be left to the players, and this is the main message of *Winning Ways*. It is in the long tradition of Rouse Ball, Dudeney and Lucas, but a great deal further along the line. Some years ago one of the authors, John Conway, published *On Numbers and Games* (Academic, 1976). The first half of that book contained a highly idiosyncratic foundation of the number-system, generating the integers, rationals and reals by a single process which yields them in other than the usual order. There was also an extension of the system to other than "usual" numbers. The second half applied this almost as an afterthought to certain games; but Conway claimed that this second half was really the main part of the book. *Winning Ways* generously redresses the balance. It is hard to characterize the content, except to say that it is mathematical games, understood as broadly as possible, and with a good deal of mathematical lessons deduced from them.

The first part of Vol.1 discusses the types of game described in Conway's earlier book. In these games two players move alternately (often from a particular starting position) according to rules allowing options, until a player loses by being unable to move. There is complete information — both players know what is happening. So this is precisely *not* the "Theory of Games", and no intervention of chance is allowed. Nim is the prototype of such games. (Elsewhere in the book some of these conditions are relaxed, but never the restriction to complete information and no chance.)

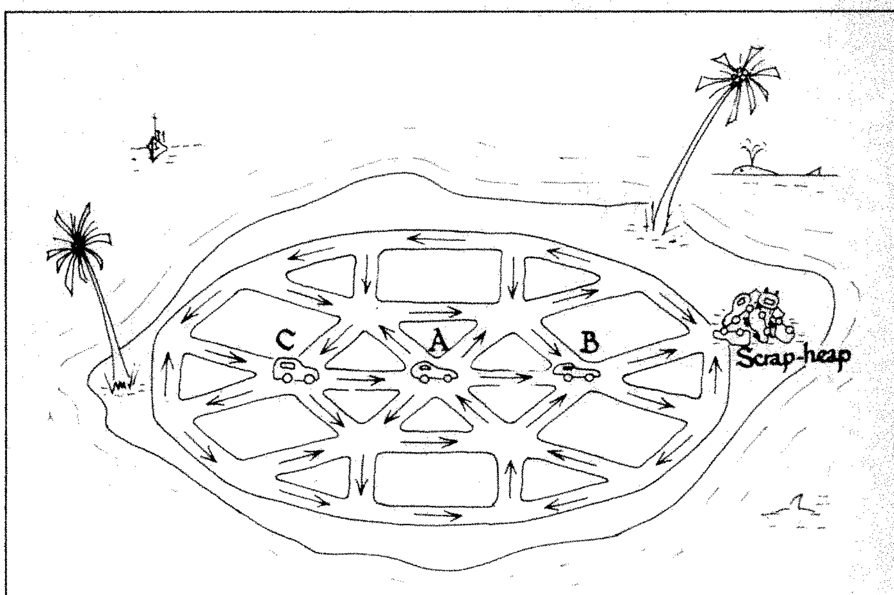
There are two foci of interest: the technique of attaching numbers to positions in games to assess their values, and the analysis of the structure of new games formed from the sums of known and analysed ones. The first of these is so simple as to excite (initial) disbelief. If in any position P the two players have options leading to positions of values, a,b,c . . ., or d,e,f . . . respectively, the value of P is written $\{a,b,c\ \ldots \mid d,e,f\ \ldots\}$. This value may or may not be a number. If a,b,c . . ., d,e,f . . . are all numbers, the value of P is the simplest number (to find out exactly what simplest means you have to buy the book), if it exists, strictly greater than a,b,c . . . and strictly less than d,e,f . . . . The beauty of this system is that it needs no start; for $\{\ \mid\ \}$, in which neither player has a move, is a zero position i.e. $\{\ \mid\ \}=0$. Then $\{0\mid\ \}=1$, $\{1\mid\ \}=2$, . . . , $\{\ \mid 0\}=-1$, . . . $\{0\mid 1\}=\frac{1}{2}$. When the value of P is a number, it has an immediate interpretation as the number of moves advantage. (Then ½ would correspond to a one move advantage in two copies of the game played together — and this leads on to the idea of the sum of two games, in which they are played side by side, and each player, when he moves, may choose to move in either game.) But even if the value is not a number, and so belongs to a natural extension of the rational numbers including infinitesimals, the ordering can be extended to say something useful.

The second part of the book, which concludes Vol. 1, goes on to games where some conditions are relaxed: you can move in several components of a compound game, or must move in all, or there are infinitely many positions, or the game may not stop. The main part of Vol. 2 applies these ideas to games of various kinds with coins, pencil-and-paper or on a board.

The approach is typified by noughts-and-crosses, of which a concise yet complete analysis is given. As the authors say:

> We have no doubt that most of our readers were bright enough as children to discover that [a tied game] always happens when the game is properly played, and only the authors of books like *Winning Ways* retain sufficient interest to study the game in any detail.

But they go on, with approval, to quote Martin Gardner's implied castigation of "many players [who] have the mistaken impression that because they are unbeatable they have nothing more to learn". These two views summarize the attitudes and lessons of the book, applied to many



An example of the games — and humour — in *Winning Ways*, taken from Vol.1. Above is a view of Corrall Island, which was a Charming Antipodean Beauty Spot before the Corrall Automotive Betterment Scheme, financed by the island's three wealthy landowners, built those expensive new super-highways. The island's two political parties (Left and Right) are now running for election and the gallop polls predict a walk-over for whichever party puts the last car on the scrap-heap. The parties have already enforced the one-way system indicated and each will alternately bribe a chauffeur to drive his car along one highway (in the proper direction) to the next intersection. You will see that there is no legal move away from the scrap-heap. How should we advise the ruling party, which must make the first move?

more complex games.

Finally, in Vol. 2, are two chapters on one-person games, and a final one on a no-person game. This is Conway's famous "Game of Life" about which a lot of recent news is here available: for example, that every sufficiently well-posed mathematical question can be converted to a question about the ultimate destinies of some Life configurations, so that Life is as insoluble as mathematics. Again, to quote: "Many computers have been programmed to play the game of Life. We shall now return the compliment by showing how to define Life patterns that can imitate computers". One of the one-person game chapters is solitaire and its extensions, the second on puzzles (the Tower of Hanoi, eight concise pages on Rubik's cube and so on).

The books are elegantly printed and written in a light-hearted vein, with a high density of puns per page, some outrageous but all appropriate. There is a wealth of illustrations — many of them are in colour — for which particular praise is deserved. A good index begins with an essential glossary of symbols.

If you make a practice of playing mathematical games this book is a must; you dare not play again without it. But even if you are not among those who indulge you ought to read it, as evidence of the insights into mathematics and physics you're missing. □

C. W. Kilmister, a Professor in the Department of Mathematics at King's College, University of London, is totally incompetent at any of the games mentioned. He has, however, used the material in Conway's earlier book to try to provide a combinatorial reformulation of quantum mechanics.

Uncovering discovery

I. Bernard Cohen

The Social Basis of Scientific Discoveries. By Augustine Brannigan. Pp.212. Hbk ISBN 0-521-23695-9; pbk ISBN 0-521-28163-6. (Cambridge University Press: 1982.) Hbk £12.50, $24.95; pbk £4.95, $9.50.

MOST of this new book by Augustine Brannigan is devoted to a review of theories about the nature of discoveries in science. The author begins with a description of various psychological theories of scientific discovery, limiting himself to the work of philosophers of science and not really investigating the accounts of psychologists. He then switches to the "emergence of a social model of discovery", which leads him into the fundamental question of heredity versus environment, nature versus nurture.

For convenience, Brannigan begins with Francis Galton's *Hereditary Genius*, intended to show the way in which extraordinary talents ran in families. In antithesis to this, he presents the work of the anthropologist A. L. Kroeber, who stressed the importance of social environment, noting how great periods of history "emerged as a function of cultural development". In particular, Kroeber emphasized the significance of multiple discoveries in science as proof of the importance of the general intellectual or cultural milieu as opposed to that of individual acts of genius. Recently the argument about multiple discovery has been advanced and developed by Robert K. Merton and his students and co-workers. Brannigan further develops Merton's ideas in a most interesting and incisive manner and then contributes a thought-provoking analysis of the work of Mendel in relation to its alleged multiple rediscovery.

This leads Brannigan into an account of the nature of discovery itself and of what he calls "folk reasoning" concerning the theory of scientific discovery. There is a concluding discussion of the sociological features of discovery, where the author develops his own views of the subject and includes a model in which "the normative considerations described by Merton" are enlarged by consideration of the "cognitive grounds" for disputes associated with multiple discoveries. The author can only be applauded for this attempt to add a new dimension to the current debate, but it is greatly to be regretted that this important idea is not developed by fully worked-out examples. And it is the same for the author's programme for analysing "the rhetorical and organizational status of methods in science". It is very frustrating that in the final page or so he no more than hints at research concerning "aspects of the scientific community, its productivity, its stratification system, and factors which influence these things". He introduces a number of suggestions, but they are shot forth with rapid fire, not illustrated or developed.

Indeed, one of the great criticisms to be levelled against this book — and against so many works of this type — is that there are passing references to many individuals and episodes in the development of science, but nothing more. Only in the case of Mendel does Brannigan attempt to do any serious history. He does, to be sure, introduce some aspects of the discovery of oxygen, but even here there is no attempt to go deeply into either the primary sources or the major scholarly investigations which could have further illuminated this topic, both for himself and the reader. As a result, the book will appeal more to those interested in theoretical questions of the sociology or philosophy of science, than to scientists or those whose primary concern is science itself. □

I. Bernard Cohen is Victor S. Thomas Professor of the History of Science at Harvard University.

Too reproductive

Colin Stern

Reproductive Immunology. Edited by Norbert Gleicher. Pp.492. ISBN 0-8451-0070-X. (Alan R. Liss/Heyden: 1981.) $86, £68.

THIS curious volume was cobbled together from some of the papers given at the first S.B. Gusberg Seminar on Reproductive Immunology, in June 1980 in New York, and from contributions solicited after the event which give it a strange chiaroscuro feel. The editor seems to have aimed at obstetricians and perinatologists, as well as immunologists and developmental biologists. There are several very good contributions but also some appalling ones. The book begins with three introductory chapters: two, on immunogenetics and immunoassay are good; the third, "Introduction to Immunology", is dreadful. It opens with a reasonable historical introduction, but the rest is full of errors of fact and misconceptions too numerous to detail. One is left asking whether this was necessary and why the author, an obstetrician whose knowledge of immunology is patently sketchy, was asked to write it.

There follow an excellent discussion of V-gene activation, solid contributions on immunoglobulin synthesis and lymphoid cell ontogeny, and a poor chapter on the development of the fetal immune system: how could the authors fail to discuss the acquisition of self-tolerance? The next part, "Pregnancy Immunology", begins with a brilliant overview of the immunobiology of the materno-fetal relationship by Rupert Billingham: clear and well-written, it makes much of the rest of the book seem redundant. There are other informative chapters here, however, on syncytiotrophoblast, uterine lymph node reactivity, maternal lymphocyte reactivity and breast milk. The next section starts well — Bill Pollack's chapter on Rhesus haemolytic disease — which makes further comment (from Cherry) unnecessary.

As is typical of the uneveness of the book as a whole, the following section contains two good contributions, on SLE in pregnancy and trophoblastic disease, as well as two frightful ones, on common denominators of pregnancy and malignancy and on the immune system in gynaecological malignancy. The last 95 pages, however, on fertility immunology, would have merited publishing on their own. All nine chapters are worth reading, particularly Bigazzi's excellent review of the immunological effects of vasectomy.

There are virtually no new data in this book. Out of 32 chapters, twelve are fair to good, nine bad to ghastly and the rest average. Reading it is likely to make the obscurity of pregnancy immunobiology darker yet, at a time when the relative importance of syncytiotrophoblast antigens, the maternal responses to them,

and the basic genetics of human pregnancy and its diseases are beginning to make a discernible pattern. Half of the contributors come from New York, which may be why some chapters have an insular feel — the authors are seemingly unaware of work going on elsewhere in the world. I cannot imagine that anyone would want to buy this book for themselves, as all the information it contains is already available in cheaper and better presented forms. □

Colin Stern is a Consultant in Paediatrics at St Thomas' Hospital, London.

Lymphokines for the nineteen eighties

Dudley C. Dumonde

The Lymphokines: Biochemistry and Biological Activity. Edited by John W. Hadden and William E. Stewart II. Pp.437. ISBN 0-89603-012-1. (Humana/Wiley: 1981.) $59.50 in the United States, $69.50 elsewhere.

IN 1982 the serious student of immunology need no longer doubt the existence of lymphokines: he has only to consult the literature. As non-antibody protein products of lymphocyte activation, lymphokines act as mediators and regulators of many aspects of lymphocyte function. After a slow start in the 1960s the pace of lymphokine research began to quicken. Semantic barriers were broken down: between mouse-doctors studying "help"; guinea-pig doctors studying "delayed-type hypersensitivity"; and human doctors studying "disease". By the end of the 1970s two international conferences on lymphokines had been held, a special review journal devoted to the subject had been launched, and lymphokine research had become one of the most vigorous fields of molecular and cellular immunology.

Now in 1982, gone are the days when sceptics viewed the term "lymphokine" as having merit only in so far as it could be used both as a noun and as an adjective. The application to lymphokine research of brighter biochemistry, supplemented by techniques of cell hybridization, cell cloning and gene cloning, has even awakened venture capital: not only is there a ferment of lymphokine literature but also a ferment of medical interest in the diagnostic and therapeutic potential of lymphokines. Accordingly the lymphokinologist of the 1980s will find that Dr Hadden and Dr Stewart have edited a book which must be compulsive reading and which must be compulsively purchased given the current pressures of information retrieval.

In *The Lymphokines* the discerning consumer will find his favourite item delicately positioned between an appetizing prologue by Byron Waksman and a satisfying epilogue by Barry Bloom. Considerable effort has been made by the contributors — all authorities — to emphasize what is known about the physicochemical nature of lymphokines and the biochemical events that attend their production and action. This theme is well illustrated in the first half of the book by chapters on chemotactic, migration-inhibitory, macrophage-activating, cytotoxic, thymocyte-stimulating and colony-stimulating lymphokines. A most useful feature is the inclusion of two chapters on the confusing subject of helper and suppressor factors (though at the time the book was finally edited, the term "interleukin" would seem to have only just been coined). There is a chapter on interferons by Dr Stewart himself; here I would like to have seen a more critical discussion of the relationships between interferons and lymphokines.

Having dealt with what we all regard as "lymphokines", the second half of the book is comprised of chapters on leukocyte transfer factor, immune RNA, macrophage secretion products including prostaglandins, monocyte/granulocyte haemopoiesis and thymus hormones. To this extent the book is heterogeneous, though the student of lymphokines cannot afford to ignore current knowledge of other factors apparently regulating lymphocyte and leukocyte function. The general standard of production is satisfactory, though the subject index lacks details and there are quite a number of typographical errors.

Yes! This is a useful addition to the lymphokine literature, although it is not necessarily the "key resource" on the subject. If the aim of the book is to relate lymphokine biochemistry to biological activity, then it must be regarded as premature, simply because much of this research remains to be done. In fact, it seems to be this very theme that runs through many of the contributions, as if the authors were aware that whilst they were writing their reviews someone else was busy cloning the mRNA of interleukin 2 in *Xenopus* oocytes, or copying the gene for γ-interferon in *E. coli*. Can one capture at any moment the pace of a fast-moving scientific field with all of its genetic, biochemical, biological and clinical implications? Drs Hadden and Stewart, together with their well-known contributors, have sallied forth where some others may have feared to tread. The result is appealing. □

Dudley C. Dumonde is Professor of Immunology and Head of the Department of Immunology at St Thomas' Hospital and Medical School, London.

Oceanography from space

John R. Apel

Spaceborne Synthetic Aperture Radar for Oceanography. Edited by Robert C. Beal, Pat S. DeLeonibus and Isadore Katz. Pp.216. ISBN 0-8018-2668-3. (The Johns Hopkins University Press: 1981.) $29.25, £17.

THE US spacecraft Seasat, launched in June 1978, carried among its instrument complement a synthetic aperture imaging radar, SAR, operating at a radar wavelength of approximately 21 cm. During the satellite's short, three-month life, this instrument provided about 1.5 million km of imagery, mapped in a 100-km-wide swath at a resolution in excess of 40 m. The high resolution imagery represents a new and important vantage point for viewing phenomena occurring near the surface of the sea. This volume, stemming from a symposium held at The Johns Hopkins University Applied Physics Laboratory in the spring of 1980, presents the first broad, coherent publication of results and understanding of radar imagery of the ocean surface.

The SAR has provided images of the sea that often exhibit puzzling and varied features which must be interpreted in terms of changes in the small-scale roughness (near 30 cm length) of the surface. The original function of the Seasat SAR was to provide imagery of long-length surface gravity waves, from which one might derive a badly needed quantity, namely two-dimensional surface wave spectra, via Fourier transform techniques. The ocean surface as seen through the radar's eye has proved to be marvellously more complicated than anticipated, however, with such additional features as internal waves, current boundaries, eddies, wind stress variations, rainfall, oil slicks and shallow bottom topographic features appearing in the radar image. While surface waves with lengths greater than perhaps 100 m can often be seen in SAR images as periodic modulations, the functional relationship between the power spectral density obtained from the film and the surface wave psd is essentially unknown. Indeed, there is much discussion in the book of the basic hydrodynamic mechanisms that allow long waves to be imaged via variations in 30 cm wave energy. It follows that the theory of the imaging process for the ocean surface is in a relatively rudimentary state.

The book opens with generalized discussions of surface wave characteristics by Phillips and Kitaigorodskii, followed by a theoretical exposition on SAR imaging mechanisms by Harger and a review of the state of our understanding of oceanic winds by Pierson. A number of more detailed research results are presented next, classified according to winds, waves and circulation. These are punctuated by dramatic examples of SAR images, most of which are of regions off the US east coast (thereby revealing the geographical orientation of the contributors to the book). The clear emphasis of the research papers is on geophysics, not radar technology. Then follow reproductions of some 21 full-swath SAR images of the western North Atlantic, New England and the central Atlantic states, prepared by the symposium organizer Robert Beal; these illustrate the myriad of land and ocean features that characterize Seasat radar images.

This volume represents the first organized exposition of the oceanic results from SAR and, as such, should be available to anyone interested in this phase of the subject. Beyond this audience, however, there should be considerable interest on the part of oceanographers, marine meteorologists, and forward-looking civilian and military users of the sea. The results presented here are so unusual and compelling that the future will almost of necessity see additional imaging radars in space; however, a great deal of research work remains to be done before maximum use can be made of such instruments.

John R. Apel, now Assistant Director for Planning in the Applied Physics Laboratory at Johns Hopkins, was until recently at the Pacific Marine Environmental Laboratory of the National Oceanic and Atmospheric Administration, Seattle.

Explosion in palaeobotany

W. G. Chaloner

Paleobotany, Paleoecology, and Evolution. Edited by Karl J. Niklas. Vol.1, pp.297, ISBN 0-03-059136-8; Vol.2, pp.279, ISBN 0-03-056656-8. (Praeger: 1981.) Vol.1 $37.50, £31.25; Vol.2 $36, £27.

PALAEOBIOLOGY has, in its own jargon, undergone a recent evolutionary explosion. Twenty years ago the study of fossil plants consisted largely of investigating their structure, and where possible reconstructing the whole plant from its variously preserved fragments. In the past two decades, however, the science has linked up with a number of bordering disciplines to encompass the development, reproductive biology, biogeography, biochemistry and a range of other aspects of past plant life. These two volumes contain a collection of papers by workers in the United States, which illustrate this diversity of new approaches. They were presented at a symposium held in Cornell University to honour Professor Harlan P. Banks on his retirement. The theme is very fitting, since over some 35 years of research and lecturing Banks has contributed uniquely to giving fossil botany an image not only of a challenging and rigorous science, but of an exciting one for students to enter.

The volumes comprise 14 articles dealing with topics ranging from Pre-Cambrian microfossils (E. S. Barghoorn) and palaeoecology (A. H. Knoll), through Devonian plant structures (C. B. Beck) and Carboniferous swamp ecology (T. L. Phillips and W. A. DiMichele) to various aspects of gymnosperm (T. N. Taylor) and flowering plant evolution (G. Retallack and D. L. Dilcher; J. A. Wolfe and W. L. Crepet). Two chapters deal largely with extant plants in the context of their evolutionary history — T. Swain and G. Cooper-Driver with biochemical strategy of primitive land plants, and R. M. Schuster with liverwort evolution. Three deal with a broader perspective of evolutionary processes, B. H. Tiffney reviewing changing diversity through the history of land plants, T. J. M. Schopf the process of genomic change, and A. M. Ziegler and others the biogeography of plants through the Palaeozoic.

Evidence for the interaction of plants and past environments is a recurrent theme. The study of microfossils is still very young, and the prospect of unravelling their ecology, novel. But Knoll is able to demonstrate convincingly the correlation between environment of deposition and various aspects (species diversity, association) of Precambrian microfossils. The occurrence of this "horizontal" variation between different assemblages, in addition to the well-documented "vertical" changes, is important for two aspects of Precambrian work: if such microfossils are to be used in dating rocks, the environmental control of assemblages must be understood; equally, if we are to interpret this early phase of plant history, it is important not to confuse migration in response to environmental change with evolutionary innovation. A different aspect of palaeoecology is reviewed by Phillips and DiMichele in a quantitative review of the vegetation of the Carboniferous coal swamps, preserved in a three-dimensional state in coal balls of the American Mid-West. These give quantitative data on biomass, species diversity and succession. Remarkably few genera constitute the dominant plants in the assemblages, in contrast to present-day humid tropical forest. Little evidence of directional succession emerges, and abiotic factors (water level, salinity) were evidently

overriding controls.

Crepet considers the reversion from insect- to wind-pollination in the catkin-bearing trees early in the history of flowering plants. This apparently occurred in the seasonally dry tropics, so that the present-day association of this group with temperate deciduous forest is misleading as a guide to its origin. Wolfe examines the Late Tertiary replacement of broad-leaved forest by conifers in the Pacific north-west. For those overwhelmed by the complex interactions of environmental factors, it is refreshing to learn that the "sole factor" in the assumption of dominance by conifers in that region was the declining summer temperature.

Perhaps the most far-reaching contribution is Schopf's reconciliation of molecular and palaeontological approaches to evolution. He concludes that "speciation may be so instantaneous geologically that many such events might well occur within the best possible resolution of the fossil record". He argues that even if our record of past life is the product of gradualistic change, the process of its formation puts its own punctuated stamp upon the picture. Which brings us, appropriately, back to Darwin's reservation over the "imperfections" of the fossil record. □

W. G. Chaloner is Professor of Botany at Bedford College, University of London.

Room for giants

Janet V. Watson

Igneous Rocks of the British Isles. Edited by D.S. Sutherland. Pp.645. ISBN 0-471-27810-6. (Wiley: 1982.) £55, $132.

FOR more than three thousand million years, melting in the outermost parts of the Earth has been an unusual event taking place only when and where temperatures rose to abnormally high levels. Igneous rocks formed by the consolidation of silicate melts therefore appear in the geological record as indicators of anomalous disturbances which are almost always related in time and space to major structural changes in the Earth's crust and mantle.

A book designed to document the record of igneous activity in the British Isles must therefore aim to throw light on the dynamic evolution of the entire crustal province if it is to be more than a source-book. From many points of view, there is a real need for a work of this kind. As Dr Sutherland notes in her editorial preface, no general synthesis has been published since Geikie's *Ancient Volcanoes of Britain* appeared in 1897. The literature which has accumulated since that date is vast, chaotic and often contradictory. Embedded in it are the outlines of fundamental ideas on the genesis, differentiation and emplacement of magmas which have become classics used all over the world. Encouraged as much by the exceptionally attractive dust-jacket as by the table of contents, geologists will therefore open this book with thankfulness and hope.

Much care has been taken to make it as useful as possible as a work of reference. The treatment is broadly chronological and each of the seven parts opens with a general scene-setting chapter. The distribution of igneous suites and the structure of individual complexes are recorded in many clear maps, well-designed for their purpose. The descriptive chapters conform to a fairly consistent pattern which makes for ease of reference. Chemical analyses, geochronological data and petrographical details are segregated in appendices (the petrographical appendix living up to a hallowed tradition that goes back to J.J.H. Teall). All these are excellent features, due, no doubt, to good planning and editing.

The contents of the descriptive chapters vary in scope and style according to the kind of information available and the predilections of individual authors. The Tertiary province comes off best, with a cluster of chapters covering the volcanics, the centres, the dykes and the petrogenetic story collectively building up a coherent picture in which details fall easily into place. Elsewhere, there are some unexpected gaps (why does the best-exposed British ophiolite, in Shetland, rate only a passing mention?) and a number of thin patches not all of which can be put down to paucity of evidence.

More disturbing, however, is the apparent shapelessness of the sections covering the central event of British geological history, the Caledonian orogenic cycle. The decision to deal separately with volcanic and "plutonic" assemblages bears especially hard on these sections where the progression of tectonic events through subduction to collision and uplift is almost lost sight of. Though defensible in terms of convenience, the separation leads to absurdities such as the allocation of the Glencoe ring complex and the ignimbrites erupted in this complex during cauldron subsidence not only to different chapters but even to different parts. The British Caledonides emerge from this book as untidily as they do in the field. One longs in vain for the authoritative touch that might have shaken this sprawling orogen into shape. The Tertiary province has never lacked giants who measured up to its challenges. There is still, on the evidence of this book, something left for giants to do in the older parts of Britain. □

Janet Watson is a Professor in the Department of Geology at Imperial College, University of London.

Peat and prehistory

A. G. Smith

The Archives of the Peat Bogs. By Sir Harry Godwin. Pp.229. ISBN 0-521-23784-X. (Cambridge University Press: 1981.) £25, $49.50.

IT HAS probably come as a surprise to many a young scientist to realize that scientific work is as much about people as it is about science. This is amply illustrated in Sir Harry Godwin's book. It is also an adventure story. Through it all runs a sense of compelling interest and excitement. Above all, however, the author illustrates the way in which ideas about the nature and importance of peat bogs have flowed down the generations of scientists. It becomes, incidentally, an interesting contribution to the history of science.

In 1935 we find Godwin confessing "my ignorance of bog ecology was vast". But he began to learn his trade in the "magical Irish countryside" (acting as chauffeur of a model T Ford) on an excursion to Ireland with Hugo Oswald and A. G. Tansley. They chanced to meet, in Roundstone, another group including Knud Jessen, the great Danish Quaternary scientist, and one of his students, G. F. Mitchell. This group had just embarked on their studies of Irish bogs. We are treated to an enduring cameo of one of the fruits of this meeting: the party slowly sinking into the mire as they listened to an apparently interminable discussion between Jessen and Oswald, in interminable Irish rain, on Von Post's concept of soligenous bog.

From this encounter came the idea among the younger workers of pooling their experience to work on an uninvestigated British peat bog. Thus in 1936 began the classical work on the stratigraphy and ecology of Tregaron Bog in Dyfed, now a National Nature Reserve, which is briefly reviewed. The story continues with Godwin's adventures in the Somerset Levels and Borth Bog (Cors Fochno), Dyfed. The first scientific investigation of this now famous bog (again a National Nature Reserve) was apparently nearly frustrated by the presence of a wide drain. Having thrown their impedimenta over it, Godwin and his companion (D. H. Valentine) had no alternative but to undress and swim for it.

It is apparent that Harry Godwin soon fell in love with British peat bogs, not only as archives, but because of their flora. This he affectionately describes, treating the various species as the *dramatis personae* of the historical pageant. In his descriptions of the bog plants he takes care to show how they can still be recognized after their preservation in the peat. These tricks of the trade are one of the many valuable side-lights of the book.

The ability to recognize the sub-fossil remains of bog plants is of course of fundamental importance in understanding

the history of the bogs themselves. The recognition of distinct layers of the remains of the giant sword sedge (*Cladium mariscus*) and their relationship to the prehistoric trackways in the peats of the Somerset Levels enabled Godwin (working initially with A. R. Clapham) to piece together their basic history. The substantial part of the book given over to this work is not only a compelling detective story: it also includes a rather charming love story for good measure! The treatment of the Somerset Levels is particularly valuable. It brings together in an easily digestible form material from a number of scattered papers. Like the work as a whole, this account can be read with equal pleasure and profit by the research student and the general reader with interests in natural history and archaeology.

The latter part of the book deals with the discovery of the effects of prehistoric farmers on the vegetation of Britain, and the impact of the new technology of radiocarbon dating on the subject. There is shorter coverage of climatic influences and sea-level changes, the origin of blanket bogs and their subsequent erosion.

Godwin's account ends intentionally at about 1960, though with most of his topics he does allude to subsequent work. How happy it is that he has been inspired to write it as a personal story. In doing so he has in no way compromised his science. The book is a pleasure to read: a book to be re-read and re-enjoyed, not only used as a work of reference. □

A. G. Smith, a former research student of Sir Harry Godwin, is Head of the Department of Plant Science at University College, Cardiff.

Scene in New Guinea with characteristic animals (from the *Geographical Distribution of Animals*, 1876), showing Wallace's selection of Papuasian animals to contrast with the Sundaic animals. Left to right and top to bottom: a tree kangaroo *Dendrolagus inustis*, the fairy lory *Charmosyna papou*, the twelve-wired bird of paradise *Seleucides melanoleuca*, the common paradise kingfisher *Tanysiptera galatea*, and a crowned pigeon *Goura cristata* (*coronata*). In the companion plate to illustrate the Sundaic fauna, Wallace chose the western tarsier, flying lemur, pentail treeshrew, Malay tapir and lesser mousedeer.

Drawing the line

Peter D. Moore

Wallace's Line and Plate Tectonics. Edited by T.C. Whitmore. Pp.91. ISBN 0-19-854545-2. (Oxford University Press: 1982.) £15, $39.

IT WAS in a letter to H.E. Bates, written in 1858 from the Malay Archipelago, that Alfred Russel Wallace first drew attention to the two distinct and rigidly circumscribed faunas which are found in that region. They are as different, claimed Wallace, as the faunas of Europe and North America, yet there is no obvious geographical reason for their separation. Often the line of demarcation passes between islands lying close together.

The cause of this biogeographic discontinuity has provided much fuel for debate and the area featured strongly in the proposals of Wegener in support of his theory of continental drift. With the modern developments in plate tectonics much of the mystery has been removed from the subject, but the area remains one of the clearest demonstrations of the biogeographical consequences of moving continents. This book is a collection of essays by various authors setting forth the current state of knowledge concerning relationships between various plant and animal distribution patterns and the tectonic development of the area.

An historical background to the subject is used as an introduction and emphasis is placed on Wallace's own indecision regarding which side of the islands of Celebes his line should lie. As a result of a succession of faunal and floral analyses during the past 120 years, the line has often migrated in the minds of men. Two essays trace the continental movements leading up to the establishment of the discontinuity, and are accompanied by some detailed cartographic reconstructions. On the basis of geological data, Audley-Charles considers that the line of convergence of the Australia–New Guinea plate with Asia actually runs through the island of Celebes, thus explaining its biogeographical complexity.

The collision was a relatively recent one, occurring about 15–5 × 10^6 years ago, but there has still been considerable time for floral and faunal migration, mixing and even evolution, and the remainder of the book consists of a series of analyses of the biogeography of specific groups of organisms (vertebrates, palms, and other plants) in the light of recent palaeoclimatic and palaeoenvironmental reconstructions.

The papers presented in this book are concise, clear, well-illustrated reviews of a variety of specialist subjects, all of which impinge upon the modern view of Wallace's line. □

Peter D. Moore is Senior Lecturer in the Department of Plant Sciences, King's College, University of London.

nature

27 May 1982

Time to chart Europe's future

Britain is once more the black sheep of Europe, after another row about agricultural policy, but the event is also an opportunity. Somebody should seize it.

This is turning out to be a bad year for the European Community. After what should have been a cheerful twenty-fifth anniversary (see *Nature* 25 March, p.278) that had most member governments asking "What have we done wrong?", a quarrel redolent of the turbulent 1960s has broken out between Britain (assisted by Denmark and Greece) and the six other states. The issue is that which has dogged the Community since it began — the price to be fixed for agricultural products in the year already under way. British unwillingness to agree to price increases exceeding 10 per cent had been well advertised in advance, as had the British government's insistence that this year, as in the past two, it should pay less towards the common budget than the rules of the Community require. With other things on its mind (see below), the British government obviously failed to sense that if it stuck to its position, the convention that any member state can veto any proposal would not save it from being voted down. This, in the event, is what happened. In the next few weeks, the member states will be looking for some way of picking up the pieces.

That will not be easy. Ever since it joined the European Community in 1972, the British government has been grumbling about the common agricultural policy, in effect a device for paying European farmers well so as to produce more food than Europe can consume. For much of that time, the British have been chipping away at the policy, and not always in the most high-minded spirit. Only a few weeks ago, a British agriculture minister was boasting to the House of Commons how he and his colleagues had managed to dip British hands into the pork barrel, winning more than a quarter of a new subsidy on suckling cows and virtually the whole of the novel subsidy for sheep. The difficulty, which successive British governments have been slow to recognize, is that the common agricultural policy, the constant but of their complaints against the Community, is as much a political necessity for other member states as is the right to veto any proposal judged to touch a vital issue is for Britain. These are eminently circumstances that call for compromise and self-restraint. Pretending that the truth is otherwise will lead nowhere. The ideal is that the common agricultural policy should wither, perhaps because it is overtaken by farmers' recognition that, in the modern industrial state that Europe seeks to be, there are more interesting and profitable things to do than till the land. What the Community should be looking for, in the rowdy weeks that lie ahead, is some means of making that ambition come nearer.

The difficulties are nicely illustrated by the grand French plan to do something within the Community about the commercial challenge from Japan in the electronics industry (see page 257). The obvious objection to the French plan for the revivification of the French electronics industry is that it is a national plan, and one based on the substantial investment of government funds. Even if the enthusiasts for the scheme were able to recruit the funds they seek, and if their use of them were wise, the result would merely be to make the electronics industry in the rest of Europe even more hard-pressed than at present. The result of that would be to renew the clamour that national governments should do even more than at present to help their own domestic manufacturers, perhaps by putting more business their way. The end result would be a system not very different from that now embodied in the common agricultural policy — a European electronics industry protected from the outside world by high tariff barriers, selling its products at prices that are too high and competing fiercely with itself. That is not a recipe for meeting the challenge from outside but, rather, a recipe for continued failure. The government of France seems not to appreciate that the Japanese success in electronics has been founded on an accurate appraisal of the international market, and that the only effective counter — admittedly only a first step — is to make the European market in electronic products thoroughly international. So will the French, and their Community partners, give up their present policies of channelling public contracts towards their domestic companies? Not, the guess is, without a struggle.

That, however, is precisely the struggle on which the British government, blackballed last week in Brussels, should now embark. The fact that the electronics industry in Europe is closely linked with defence, specifically exempted by the Treaty of Rome from the rigours of the open market, is not the stumbling block it seems but rather the opposite. It is foolish that half of the ten governments in the European Community should be spending large sums of money on the solution of essentially similar problems, hoping to recover part of the cost by sales of advanced military equipment overseas. Why should not each of them spend less on research and development, selling the products to each other in the first place? The result could well be that member states' strategic independence would be compromised, but is it seriously suggested that the constant rows within the European Community imply that one day it may be necessary for the members to attack each other? The pretence of military independence in which the member governments of the European Community at present luxuriate is belied by what their electors know — that culturally the Community is already so much of a piece that unscrambling it would be literally unfeasible. The advantage to the British government of taking the initiative on this issue of electronics and, where necessary, defence, is that it could throw those who baited it last week onto the defensive, yet hope to emerge with a prize to offer its own people that would be at least as valuable as the cost of the agricultural policy.

Rules for limited war

The Falkland Islands conflict marks a step towards the automation of conventional war.

The politics of the Falkland Islands war is outside *Nature*'s bailiwick, but the technology is not. For what is happening now in the South Atlantic is probably a foretaste of what may happen all too often in decades to come. First, it is a limited war in which neither side seeks the comprehensive capitulation of its adversary. Second, it is a high-technology war in which the combatants are equipped with virtually automatic weapons developed against the possibility of conventional conflict in Europe, but now being used in a mercifully remote part of the world, one so distant that other states are unlikely to be drawn in. What lessons, spine-chilling and otherwise, can be drawn from what is now going on?

That limited wars, out of fashion since the eighteenth century, are still feasible should not be a surprise. The Yom Kippur war of 1973 was limited in the sense that the Israeli objective was the capture of a tract of territory, the Sinai desert up to the Suez

0028-0836/82/210253-02$01.00

Canal, and not the destruction of Cairo or of the Egyptian state. This self-imposed restraint was not based on high-minded principle but on a calculation of what might be accomplished without dangerously broadening the conflict, reinforced towards the end by the discovery that Egypt was well-equipped with surface-to-air missiles. (We are all lucky that the calculation was not disproved.) Similarly, the involvement of the United States in the long war in Vietnam was limited; it was tacitly understood that nuclear weapons could not be used without triggering off a broader conflict that the United States did not seek. The same is true in the Falkland Islands. The United Kingdom has a modest stock of nuclear weapons, some conveniently carried on submarines, but their use is unthinkable and has not been mentioned. Even the wild talk in Britain that it might be necessary to attack airfields in Argentina has quietened as the sober realities of conflict have dawned, and as the importance of public and political opinion elsewhere in South America has been appreciated.

What this implies is that limited wars are possible only with the tacit agreement, however grudging, of other interested parties. And because other states' opinions can change, the combatants must spend a substantial part of their energy persuading everybody who is prepared to listen that they are acting reasonably, or at least as reasonably as is consistent with their resort to violence. This, no doubt, explains why the Falklands conflict has so far been conducted with extraordinary civility. The substantial British community in Argentina has not been unduly harassed, while in both capitals journalists go normally about their business (although three British journalists have been wrongly locked up in Argentina for six weeks, and three others kidnapped for a day). In Britain, the only untoward incident has been the decision last weekend by the Tottenham Hotspur football club that its talented Argentine player should be dropped from Saturday's cup final. Both sides have an interest in continuing like this, for to behave otherwise would weaken the cases they are putting to their friends.

Against this background, the British government has made two important errors. First, it has unnecessarily sought to keep to itself news of what is actually happening in the Falkland Islands. Second, no member of the British government has yet made the conciliatory speech that circumstances demand, making the simple point that the traditional links between Britain and Argentina need not be jeopardized by what is going on and even that the present Argentine government, while not democratically elected, is less fiercely repressive than its predecessor. Indeed, it would be entirely consistent with the notion that the Falklands war is strictly limited that even now the British government should be prepared to negotiate the kind of settlement it will be prepared ultimately to countenance. For the time being, both sides are stuck on the abstract concept of sovereignty; sooner or later, they will have to settle among themselves some kind of monetary price at which that concept could be made more flexible (see *Nature* 8 April, p.480). One of the tricks in fighting a limited war is to help opponents to give up the struggle at the earliest opportunity.

Technologically the Falklands war is far from limited. Reports that troops under air attack are unnerved not to see the attacking aircraft but only the missiles sent automatically to bring them down are vivid and credible, but that is what conventional warfare has become — a technicians' war. The past few weeks have shown that vehicles carrying people, whether ships or aircraft, moving in an environment in which radar detection at long range is technically straightforward, have become exceedingly vulnerable to attack. As missile ranges lengthen, they will become more so — which should give pause to the now-vociferous British naval lobby. Long-range attacks on ground targets are for the time being more difficult and still require human intervention, but the development of conventional battlefield weapons in the past decade suggests that even that relative immunity will not persist indefinitely. It does not, of course, follow that people will cease to matter. On the contrary, their roles as gatherers of intelligence and as communications links between pieces of equipment will become more important while their numbers shrink and as their skills increase. This inexorable trend points to several circumstances. First, limited war becomes more feasible as the chances of large numbers of military personnel being killed diminish. Second, the cost of military preparations is bound to increase. Third, there is a limit to the extent to which the depersonalization of conventional warfare can be carried without the combatants deciding that they must force a decision by using civilians as hostages of a kind. None of these prospects is cheerful.

Rothschild rides again

Lord Rothschild's report on social science research is predictably intelligent and subversive.

Once upon a time, the story goes, there was a wicked fairy called Sir Keith Joseph, Secretary of State at the Department of Education and Science in the United Kingdom. One day last year, soon after his appointment (some say demotion) to that office, Sir Keith took the daring step — it was at least unprecedented — of cutting half a million pounds from the recommended budget of an innocent but youthful mainstay of British academic life, called the Social Science Research Council or SSRC for short. Anxious to dissemble the extent of his wickedness, Sir Keith said that he would leave it to Lord Rothschild, a scientist and merchant banker who acquired in 1971 a fearsome reputation for angering academics, to recommend whether or not this stripling outfit should continue to exist. Thanks to the journal *New Society,* it soon became known that Sir Keith had been dutifully promising the Chancellor of the Exchequer (by definition wicked) that *something* would be done about SSRC. Thunder and lightning occurred.

Wicked fairies have no luck. How could Sir Keith have known that his chosen companion in malevolence had earned his reputation for wickedness by his flair for telling the truth and doing it deftly? Lord Rothschild's report on the issue last week (*An Enquiry into the Social Science Research Council,* Cmnd 8554, HMSO, £6.50) is in that tradition. For as nice a statement of the case for social science, and a neat dig at Sir Keith in the process, what could better the fifth paragraph of his report?

> Einstein is reported to have said that "Common sense is a deposit of prejudice laid down in the mind before the age of eighteen". Yet to the layman — the man in the street — many of the problems on which social scientists work are either too far fetched or do not need further work because common sense answers are already available. Common sense tells us that:
> (1) Capital punishment deters potential murderers;
> (2) The import of foreign cars should be drastically curtailed;
> (3) There should be more small hospitals and fewer large ones;
> (4) Horror comics make children violent (USA).
> Are we sure that common sense is right? Might there not be a layer of prejudice or even ignorance?

So, this is how the story goes, Lord Rothschild turns out not to be a wicked fairy's hatchet-man, but a good fairy in disguise. The youthful SSRC has been misjudged, he says. Too many have expected too much of it too soon, and without a proper understanding of what academic life or the social sciences are like. So the SSRC should be reprieved, first from the burden of being "looked at" for at least the next three years. In gratitude, however, it should be prepared to move to Swindon, should be more (not less) courageous and should defend itself against the single open charge of left-wing bias. Wicked fairies everywhere will be discomfitted by this sane and unchallengable opinion. One of these will be the British Academy, which calls itself "the national learned society representing the interests of the social sciences", and which took the lead in bad-mouthing the SSRC a year ago. Lord Rothschild (who is evidently not unalloyed good fairy) quotes a pompous passage from the academy's written evidence saying (truthfully) that it is incompetent to be the SSRC but promising that it will be "vigilant" in overseeing the council's work. Can everybody live happily after that?

0028-0836/82/210254-01$01.00

UK plutonium worries US Senate

Alarm at link between civil and bomb uses

Washington

US senators may soon start investigating the fate of as much as four tons of fuel-grade plutonium that the United Kingdom supplied to the United States over several years on the understanding that it would be used only for civilian purposes, but which may have been put into nuclear warheads by US authorities. The plutonium originated in the Magnox reactors of the Central Electricity Generating Board.

Last October the two governments apparently agreed that the United Kingdom would resume providing plutonium to the US Department of Energy (DoE) to help indirectly with the DoE's programme for producing and modernizing US nuclear warheads.

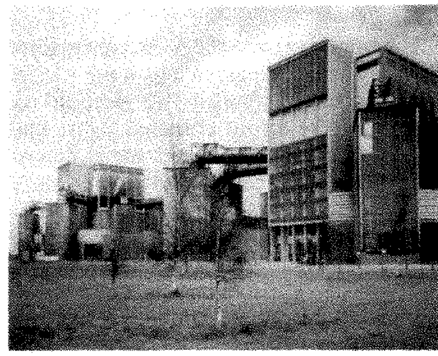
Under a previous barter agreement that ran from 1964 to 1971, the United States supplied the United Kingdom with highly enriched uranium for its nuclear submarine programme and with tritium for UK nuclear weapons. In return, the United Kingdom supplied an unspecified amount of plutonium from Magnox reactors. The terms of the accord have always been secret, but in 1964 a British government spokesman indicated that the United States had no intention of using the plutonium thus obtained for nuclear weapons. Now, however, Senators Gary W. Hart (Democrat, Colorado), Alan K. Simpson (Republican, Wyoming) and George J. Mitchell (Democrat, Maine) seem likely to act on a request from a public interest group that they look into how much plutonium was shipped to the United States between 1964 and 1971.

The Senate investigation reflects a growing concern in the United States about DoE's plans to link civilian and military nuclear programmes, a move away from the historic separation between "atoms for peace" and "atoms for war". Recently, DoE has announced that there is a "shortage" of plutonium (DoE designs and builds nuclear weapons for the Department of Defense). So it plans to take plutonium from spent fuel from several civil research reactors to fashion more weapons grade material.

DoE also wants to convert the 17.8 tonne stockpile of fuel-grade plutonium now destined for the breeder reactor at Clinch River, Tennessee, for weapons use. This stockpile may well include much or all of the plutonium supplied under the barter agreement. "Either it is in the US nuclear weapons already, or it is in the stockpile meant to fuel the breeder, but now destined for weapons," says S. Jacob Scherr, senior staff attorney for the Natural Resources Defense Council Inc.

According to the Natural Resources Defense Council, existing US nuclear weapons utilize about 90 tonnes of plutonium (plus or minus 15 tonnes). Current weapons-grade plutonium production by DoE has run about 1.4 tonnes per year, but DoE wants to step this up to 2 to 3.5 tonnes annually. In October, the UK government allegedly agreed to send more plutonium to the United States. The British plutonium may be going to fuel the breeder at Clinch River so that fuel otherwise destined for the breeder can be diverted into nuclear weapons. John Moore, Parliamentary Under-Secretary of State at the UK Department of Energy, was questioned about the plan in March.

DoE seems likely to continue to press the United Kingdom for more plutonium, because Senators Hart, Simpson and Mitchell have succeeded in cutting off another possible source — by a vote of 88 to 9, on 30 March, the Senate voted for an amendment to prohibit DoE from using spent fuel from commercial reactors for its nuclear weapons programme. In a debate in the House of Commons on 21 December 1981, Mr Moore emphasized that civil plutonian sold to the United States would be subject to international safeguards, but he implied that military plutonium transferred under the defence agreement would not be so restricted.



The Magnox at Berkeley in Gloucestershire — source for US bombs?

DoE has also decided to proceed with the Laser Isotope Separation facility at Lawrence Livermore Laboratory, which would cost an estimated $560 million and which the Natural Resources Defense Council claims has no justification other than to enrich plutonium from civil reactors to weapons-grade material. DoE also seeks to build a decladding plant to utilize fuel from noncommercial plants, such as the Fast Flux Test Facility at Hanford, Washington.

The issue at stake is not only the historic separation between civilian and military nuclear programmes, but also the precedent that DoE is setting for other countries. Critics of DoE's plans argue that if the United States starts taking civilian spent fuel from its weapons programme, it will not have a leg to stand on when it tries to dissuade other countries from doing likewise.

Deborah Shapley

Oak Ridge up for grabs

Washington

On 3 May, the Union Carbide Corporation, which has run government atomic facilities at Oak Ridge, Tennessee, since the Second World War, announced that it no longer wants to do so. Union Carbide is not seeking to renew its option to manage Oak Ridge when the present contract expires on 30 September 1983. Recognizing the dislocation that its decision could cause, however, the company says it will stay on for up to three years longer if necessary.

The names Oak Ridge and Union Carbide have been almost synonymous in government science circles and in the Tennessee Valley region since the first gaseous diffusion plant was built there, amid great secrecy, during the Second World War. At that time, the company was heavily involved in manufacturing carbon products, but has since grown (its 1981 sales were $10,000 million) and diversified into batteries, antifreeze, car care products, even salmon farming and ferroalloys. But earnings have been down of late, and the company has been divesting itself of activities which are not central to its purpose. So it is shedding Oak Ridge.

Union Carbide is one of the most respected corporate managers of any government laboratory in the nation. It actually runs four facilities — Oak Ridge National Laboratory, the Oak Ridge gaseous diffusion plant, the Oak Ridge Y-12 plant that makes nuclear weapons components, and the gaseous diffusion plant in Paducah, Kentucky. Union Carbide's act will be hard to follow.

The government spends about $1,000 million a year on these facilities, paid through the company, which has 18,000 employees working there on its payroll. Its fee for the operation is $8 million per year, from which come some expenses. The Department of Energy, which holds the contract with Union Carbide, may decide to open four separate bids, one for each of the facilities, rather than ask a single company to follow in Union Carbide's footsteps. As a side effect, this might hasten changes in the mission and role of the Oak Ridge laboratory itself, which is under White House review, along with the other departmental laboratories.

Deborah Shapley

0028-0836/82/210255-01$01.00

Hamburg elections

Power politics

Heidelberg

Nuclear power and the environmentalists will play a crucial part in next week's local elections in Hamburg. The election on 6 June will be a four-cornered fight. The looseness of the traditional social democratic grip on the Hanseatic city state reflects the party's national difficulties. Mayor Klaus von Dohnanyi may well be returned, but with a minority in the senate. Proposals for a *gross Koalition* between the Christian Democratic Union (CDU), the second largest party, and SPD have been dropped. The Free Democrats (FDP), aligned to SPD, are fighting for the 5 per cent vote needed for survival under the proportional representation rules. But the crucial minority, likely to get 8–10 per cent of the vote, is the Grün-Alternative Liste (GAL), an umbrella party of some 150 groups based on issues ranging from citizens' protection and women's rights to environmentalism and opposition to nuclear power.

Mayor Dohnanyi would apparently be willing to attempt a minority administration with GAL cooperation. Despite the difficulties of arriving at even an internal GAL consensus, and despite its preference for a role in opposition, GAL may agree. But its terms include rejection of the new *Energiekonzept*, the controversial product of long hard bargaining between Hamburg and the Hamburgischen Electrizitätswerke (HEW).

These plans, announced on 27 April and heralded as classic reforms, involve going ahead with what is probably the country's most disputed nuclear reactor, at Brokdorf, and also the construction of four conventional coal-fired power stations with heat-coupling. Under the previous plan, 70 per cent of Hamburg's electricity was to have come from nuclear reactors; this has now been reduced to 25 per cent with the possibility of extra conventional capacity that would allow Brokdorf to be abandoned by the year 2000. HEW will invest DM3,000 million for the electricity-heat project and DM2,500 million for Brokdorf. DM50 million will be contributed by the federal government, which sees the plan as a pilot project for integrated energy supplies in built-up areas. A third of a million homes will be heated, 1,500 jobs created, 140,000 tonnes of oil saved, and pollutant emissions significantly reduced.

The plan provides ammunition all round. GAL insists on withdrawal from nuclear power plants at Brokdorf, Brunsbuttel and Krümmel. Nor is the prospective advent of a friendly neighbourhood coal-fired power station, however smart, pleasing to that diverse conglomeration. Within SPD, nuclear power is strongly supported by the federal government but rejected by both Social and Free Democratic Parties in the north. The local Hamburg SPD voted last year for withdrawal from Brokdorf and the previous mayor, Hans-Ulrich Klose, resigned on the issue. The new plans split Dohnanyi's home base and may well drive SPD left, nearer to the "Greens".

Despite Dohnanyi's sleight of hand, his attempt to back the nuclear power issue both ways may leave him stranded. The Christian Democratic mayoral candidate, Walther Leisler Kiep, has attacked the financial basis of the project and the move from clean environmentally favourable nuclear energy to coal, predicting higher energy prices and an electricity gap if local objections slow construction of the coal-fired power plants.

Peter Glotz, party executive secretary, compares the social democratic response to the unwieldy manoeuvres of a tanker. If Hamburg's "Greens" reach the bridge, the Dohnanyi *Energiepolitik* may go overboard and the national supertanker be on course for a dry dock. **Sarah Tooze**

East–West energy trade

Pipeline ploy

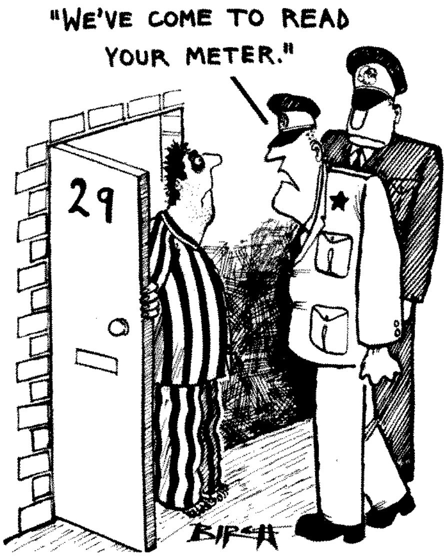
President Reagan's visit to Europe next month will include major talks on the future of East–West trade. The immediate issue is that of the interest rates to be demanded of the Soviet Union. The Americans' wish to increase them is opposed by some West European countries, notably West Germany, which feel that if the Soviet Union is reassessed, several other countries should be treated likewise.

The Soviet Union has a chronic shortage of hard currency. One possible means of repaying loans would be a counter-trade in fuel, particularly natural gas, which during the 1980s will increasingly replace oil as the Soviet Union's major fuel export. French and West German contracts for deliveries of gas from the Urengoi field originally caused the Americans considerable concern. Of late, however, they appear to have accepted European arguments that the amounts involved (10,500 million cubic metres for West Germany plus 700 million for West Berlin and 8,000 million for France) form only a small proportion of the energy budget, and do not constitute dependence. However, the signing of a contract last week for the supply to Switzerland of 360 million cubic metres of gas annually for 30 years suggests that other West European customers for Soviet gas might not follow the comparatively small demands of France and West Germany.

The transfer of technology, heavy-duty pipelines and compressor equipment, as part of the initial Western investment in the pipeline, continues to worry the Americans, although little or nothing is involved that had not been made available to the Soviet Union ten years ago.

A recent survey of East–West energy trade, however, suggests that the lack of a coherent Western policy could considerably weaken the position of the NATO countries in the changing energy situation of the 1980s. The survey's author, Jonathan Stern, describes a dual energy scenario for the 1980s.

On the one hand, Stern says, an endemic energy shortage will develop in Eastern Europe which will only be made good by purchases on the hard-currency market. Under present agreements, the Soviet Union cushions increases in the price of oil and gas to the Comecon countries, but only up to an agreed amount, leaving an increasing gap to be filled from the world market. Unless the Comecon countries can agree with the oil producers to purchase oil and gas with soft currency, foreign debts will escalate and political and economic tensions will increase.



On the other hand, gas exports are an attractive hard-currency earner to the Soviet Union. Construction of the pipeline, in spite of the delays imposed by the US sanctions after the military takeover in Poland, is likely to go ahead, and once the pipeline is in operation, further contracts with the West are likely.

A coherent Western policy, Mr Stern urges, should therefore be worked out, based on the following criteria:

- Are other gas suppliers a more acceptable security risk?
- Are the security risks outweighed by related contracts?
- What would the Soviet Union achieve by cutting supplies (granted that it would lose more than 50 per cent of its hard currency earnings by doing so)?

Lack of such a policy, Mr Stern argues, would not only allow the Soviet Union to claim a political victory (by pointing out cracks in the Western alliance), but would also weaken the Western bargaining position in dealing with the other aspect of the East European energy problem — the efforts of the energy-poor countries of Eastern Europe to top up supplies on the world market. **Vera Rich**

0028-0836/82/210256-01$01.00

Information technology

Send more bits

The age of electronic mail is about to dawn in Europe, not traditionally the leader in new information and communications technologies. Before the year is out, most European telecommunications administrations are expected to introduce a new international standard that will allow word processors to communicate with each other through existing telephone and data transmission networks.

With its penchant for confusing names, the information technology community is calling the new service "teletex". It is, however, a far cry from teletext, the broadcast information service. Teletex is either a sophisticated form of telex or a standard that allows computer terminals to communicate with each other.

The teletex standard was adopted by CCITT, the international telecommunications standards agency, at the end of last year. Europe's enthusiasm for it stems from the present difficulty of linking incompatible word processors over telecommunications links. The United States, which might be expected to be the largest market, is reluctant to adopt the standard, chiefly because the already-large word processor market makes it worthwhile for manufacturers to adopt their own communications standards.

Theoretically, it should be possible to write teletex-compatible communications software for any type of word processor. The practical difficulties, however, have persuaded manufacturers to design specific teletex machines. Among the leaders are Siemens of West Germany, which already produces teletex equipment compatible with the German data network. Other companies, either selling teletex machines in some countries or promising them soon, include Philips, ITT Creed and IBM.

Teletex's chief attraction is that, in many instances, it can deliver documents much faster and more cheaply than the ordinary telex or the post. As it is basically a word processor, it can also relay information in different formats, unlike telex. Most telecommunications administrations will be introducing services with transmission rates of 1.2 kbits per second, doubling this later on. Using the lower data rate, a typical A4 page of text (*Nature*'s page size) will take roughly 30 seconds to transmit.

The cost of using the service will depend on the charging policy of the different telecommunications administrations. British Telecom, for example, will be making a "small" service charge but will thereafter charge for time at the ordinary tariff. At current rates, one A4 page would cost 12.9 pence to send a distance of more than 56 kilometres between 8 a.m. and 1 p.m. over the public telephone network. Forty-three A4 pages, however, could be sent within the same city after 6 p.m. for only 4.3 pence. These charges should be compared with the cost of a first class stamp, now 15.5 pence.

The capital cost of equipment is still uncertain, but manufacturers expect teletex machines to cost no more than sophisticated word processors.

The service to be launched throughout Europe this year will be relatively simple, but a facsimile standard is expected later. British Telecom's service will initially be suitable only for communicating between teletex machines through either the public telephone or packet-switched data networks. However, the service is expected to expand in 1983 to allow communications from teletex to telex machines.

West Germany, which introduced its own teletex standard in part compatible with the international standard, has already launched a service. Manufacturers that have supplied to the West German market, however, may not be able to sell their equipment in other countries without first modifying it for other networks.

US companies showing an interest in teletex seem to be those with a strong interest in Europe. IBM is planning a teletex version of its display writer word processor and Western Union is also apparently developing teletex. Canada, however, is not showing its neighbour's reluctance. It also plans to introduce a service this year. **Judy Redfearn**

French electronics

Rushing ahead

Has the French electronics plan come too late? M. Abel Farnoux and his team at the Ministry of Research and Technology in Paris have laboured for eight months to produce a grand plan for the whole French electronics industry, from primary materials right through to consumer products and computer components. Its coherence and grandeur are marvellous to behold, but news is still awaited from the Council of Ministers about the scale on which it will be financed. Twelve months ago the new government was full of expensive plans; more recently, the finance minister reckoned up the bill.

Nevertheless, Farnoux and his chief — science minister Jean-Pierre Chevènement — are pushing the plan strongly. The plan must be implemented "within a year" to be effective, they say.

Alongside industrial expansion, it would increase research and development in electronics and related fields by two-thirds, relative to the 1980 figure of FF 12,000 million, to FF 20,000 million in 1986. The strategy would cover the whole industry and not just the present strong points of telecommunications and telematics, and it would be aimed at bringing France closer to the levels of the United States and Japan. (Electronics for the French defence industry, third in the world's armaments league after the United States and Soviet Union, is also included in the strategy but has been omitted from published versions for security reasons.)

Nevertheless the published report says that electronics is a "basic technology at the heart of our defence and communications systems". Strength in electronics would also permit "indispensable gains in productivity" in other industries; open up vast consumer markets; contribute to energy conservation in new energy technologies; directly and indirectly reduce the French trade deficit; and (a point added for Mitterrand?) play a major role in "cultural" affairs.

The industry (or *filière*, as the French call a collection of industries taking a raw material to a product) must be treated as a whole, says Farnoux; for him it is all or nothing. Japan and the United States must be met head-on: it is no answer to manufacture American and Japanese products under licence, says the report. European cooperation should be sought, particularly in the development of new micro-electronic consumer products.

Division of the world electronics market, according to the recent French report

Country	Production ($\$10^9$)	% Of world production	Trade balance ($\$10^9$)
USA	134	46	+4
Japan	46	16	+13
West Europe	76	26	−6
(Germany	23	8	0)
(France	17	6	−0.2)
(UK	15	5	−0.2)
Others	35	12	−11

In France at least, a big training programme will be necessary: up to 1990 the report envisages the creation of 75,000 "engineer-researchers", 25,000 senior technicians and 400,000 skilled workers.

If the report were adopted in full, there would be 14 national projects, including a large French scientific and industrial computer, chip manufacture, electronic publishing, computer translation and others. M. Chevènement has announced the setting up of working parties on each project, independently of the report's adoption by the Council of Ministers.

On research, new units would be set up combining both academic and industrial interests. Government investment in research would rise (50 per cent of the French electronics industry is now nationalized, with 30 per cent in American hands and the rest in private French ownership) but mechanisms would have to be found to give a better return on each franc. The weak areas in current research are seen to be information and communication science, solid-state physics and chemistry. Mathematics and robotics are considered well advanced.

One final recommendation may not be well received by M. Chevènement, however. After a long exercise in gathering the instruments of technology politics under one roof, the minister is now faced by a proposal to create a secretary of state for electronics. **Robert Walgate**

0028-0836/82/210257-01$01.00

Science in US Congress

Fair winds

Washington

The apparently pro-science and technology mood of Congress this session, particularly in matters involving private industry, is indicated in much of the routine business of bills introduced, of bills reported out of subcommittee (which means that a small number of congressmen agree something should be done), of bills reported out of full committees (which means that a larger group of congressmen agrees), and by bills actually voted by either the House or the Senate.

True, Congress is not rushing through sweeping reforms. Indeed, it is unlikely to do much — in terms of legislation passed by both houses that then becomes law — that will change the face of US science. The Republican party dominates the Senate; Democrats have a majority in the House. Democrats are preoccupied with the budget battle, while the Republicans tend to be thinking of the next election.

One measure with a chance of passing is a patent reform bill (the Uniform Science and Technology Research and Development Utilization Act or S. 1657 in the Senate). This would extend to most organizations performing government research the patent reforms enacted last year for small businesses, non-profit institutions and universities. It would also unify the patent policies of the various government agencies. The long-standing question has been when a researcher using government funds is entitled to hold the patents arising from the work, or when patent rights should go to the government department that sponsored the work. The present congressional mood includes greater consensus that federal shackles be removed, allowing researchers the greatest incentive to market their products.

Another bill, the Joint Research and Development Act (HR. 6262 in the House), is a response to the US high-technology industry's complaint that other countries allow industries to pool talent on research problems but that, in the United States, such pooling risks violating antitrust laws. The bill would allow the government's lawyers to issue a certificate permitting joint research and development in selected cases, and protecting the companies from antitrust prosecution.

The Senate has passed the Patent Term Restoration Act (S.255) whose counterpart is now in the House Judiciary Committee, but may not emerge before Congress adjourns in September. It tries to help industries that the government regulates to recoup more money from patents, to compensate for the costs of regulation. At present pharmaceutical companies file for a patent as soon as a new compound is discovered. The patent runs for 17 years, of which several are used to develop the compound into a marketable drug. Then the firm must file for permission to market from the Food and Drug Administration. By the time the drug is approved for marketing, which can take up to 10 years, the company has only a few years left in which to recoup its investment. The pharmaceutical industry claims that this delay can cost $70 million for a single drug. The new bill would extend the lifetimes of certain patents by up to 7 years.

In the Senate especially, legislators seem concerned with freeing industry and assisting US high-technology trade. A resolution has been introduced to guide the imminent talks in Geneva concerning the General Agreement on Tariffs and Trade (GATT), that refers specifically to US high-technology trade needs. Another bill, passed by the Senate, would fund a special clearing-house to help move the government's enormous store of technical information into the private sector. A further measure introduced in both houses would offer tax credits to manufacturers of computing equipment which give hardware to schools. The measure was promoted by one of the founders of the US home computer company, Apple Computers.

If sentiment were more like that of ten years ago, when faith in federal government intervention was far stronger, Congress might now be designing large federal programmes to "rescue" the US high-technology industry, or greatly increasing spending on federal research and development. Instead, there is a feeling that government is not very good at picking winners and that the congressmen want to encourage promotion of technology in the marketplace. This attitude, particularly prominent among Republican senators, is in contrast to past enthusiasm for heavy federal involvement and big government development and demonstration programmes. Basic research has benefitted from the change — both those who favour more federal intervention and those wanting to promote technology in the marketplace view basic research as an essential government investment.

Likewise, the cause of improving US science education in the schools has support from both sides. Senator John Glenn (Democrat), the former astronaut who has made science and technology a main plank of his political activities, has introduced a bill (S.2421) to set up a council in the National Science Foundation to suggest a cure for the "technological illiteracy" of the nation. It would be given $5 million to come up with the plan, and $50 million per year for four years to implement it. A similar bill has been introduced in the House by Don Fuqua (Democrat) and Doug Walgren (Republican). Neither bill is likely to get very far. But the momentum these congressmen are giving to the issue of science education may promote a change of heart from the Reagan Administration, until now opposed to a major federal role in science education. **Deborah Shapley**

Computers for free

The US computer industry is joining the ranks of those crying for improvements in education in science and engineering offered in US schools and colleges. As a result, the National Science Foundation (NSF) is expected to announce in early June that five computer companies will be donating many hundreds of individual computers to help solve the growing problem of "technological illiteracy".

It all began when two computer companies — as yet unnamed — each tried to donate 100 machines to NSF for distribution to schools. This gift, however, set NSF bureaucrats worrying whether it was legal to accept this largesse. As it turned out, NSF, unlike some other government agencies, has specific statutory authority to accept gifts that are for the purpose of furthering NSF's missions.

But NSF did not want to be seen to favour these two computer companies over any rivals for the honour of giving away their machines to the government. So they went through a moneyless bidding process, and invited gifts from all companies. Now, NSF sources say, five companies will be making the donations, although the terms, the nature of the hardware, and the insitutions they will be given to have not yet been revealed.

Why is the computer industry so eager to provide free samples to young people in the schools and colleges? One answer, of course, is that a student who learns an elementary computer tongue at school will outgrow it and ask for another model. Company sales would not be hurt. **Deborah Shapley**

British universities

More misery

Hopes that the British university system would be spared some of the government's economy measures were dashed last week, when the University Grants Committee made public the recurrent grants to individual universities for the academic year 1982–83. There is no substantial change from the provisional allocations of a year ago, although the University of Salford, one of the most seriously afflicted then, has been given an extra year in which to reduce its establishment.

The coming academic year will be the second of the three in which government subvention for the universities is to be reduced by 8.5 per cent. The sum now offered to the universities is, however, larger than the amount advertised last year because allowance has been made for inflation (4 per cent on salaries, 9 per cent on other costs) and because the University Grants Committee has been given more

0028-0836/82/210258-02$01.00

than £100 million extra to compensate universities for the reduction of fees for home students.

For many British universities and academics, last week's announcement will seem to bear directly on the pay negotiations now under way between the universities and university teachers, represented by the Association of University Teachers. For the grants committee's letter does formally confirm the UK government's intention that the university grant for the coming year should include only 3 per cent for salary increases. While there have been some suggestions, at the University of Aberdeen for example, that academics might forgo pay increases in the present round of pay negotiations, the union nationally is asking for 14 per cent, 12 per cent to compensate for price inflation in the past year and 2 per cent to make good the erosion of academic pay.

The allocation of funds for the coming academic year has apparently been made in the light of universities' accounts of how they plan to adjust to falling budgets. The grants committee is apparently planning to keep back some £20–30 million of the total government grant to finance the more interesting of universities' intended innovations. The new letter to universities pleads, however, that universities having to reduce costs should not take the knife to easily eliminated but academically important minority departments.

The committee has also shared out among its dependent universities the annual government grant for equipment and furniture, fixed last month at £83.6 million. While the Department of Education and Science said last week that the grant is "consistent with the aim of maintaining standards . . .", the grants committee seems strongly to hold that the grant is at least one third too small.

For the more distant future, the grants committee seems to expect that there will be a return to "level funding" after the present contraction is over in 1983–84, but does not know whether the provisional budget for 1984–85 published in the government's expenditure white paper in March will be adjusted upwards if inflation exceeds the supposed 5 per cent a year.

The grants committee itself plans to spend much of the coming year studying possible changes in the social function of universities, especially in continuing education. It remains unclear what will befall those universities which fail to meet the grants committee's targets for reduced student numbers by the end of 1983–84. The sentence in last year's letter suggesting that universities failing to meet their targets would be penalized is not echoed in those sent out last week, but the committee is apparently guessing that if the government should be disappointed with the universities' performance, and should cut the total budget by the extra cost of student maintenance, the budgets of the universities responsible will also be cut.

London medical teaching

Merger fever

While the rest of the University of London continues to agonize about its future, the undergraduate medical schools are at last beginning to implement a plan designed to save 10 per cent of the university's medical education bill by 1983–84. The outlook for dental education also looks brighter since the five dental deans recently agreed to move the school at the Royal Dental Hospital from its premises in Leicester Square to one of the university's four other dental schools. But the future of the postgraduate medical institutes, whose finances have been particularly badly hit by the shortfall in the number of overseas students, will not be tackled until July.

After almost two years of often bitter wrangling, the undergraduate medical schools, collectively the largest source of trained physicians in the United Kingdom, finally agreed on a plan at the end of last year. The dispute began after a committee chaired by Lord Flowers, rector of Imperial College, recommended at the beginning of 1980 that the university's 34 medical and dental institutions should amalgamate into six large conglomerates.

That debate was overtaken by events. The government's announcement at the end of 1980 of large cuts in university education concentrated minds. The plan adopted was designed to increase opportunities for pre-clinical students to choose between medical and multi-faculty schools, and provide access to certain major disciplines, clinical pharmacology, therapeutics and community medicine.

Three of the medical schools, those at the Royal Free, St George's and St Mary's hospitals, are to remain much as they are. The others are to form some type of association either with another medical school or with a multi-faculty college. Thus the medical schools of Charing Cross and Westminster hospitals are to merge. So too are those at the Middlesex Hospital and University College. The schools at St Thomas's and Guy's hospitals are to form a united medical school next autumn and those at St Bartholomew's and the London hospitals are to work towards formation of a joint school. The school at King's College Hospital is to be administered from King's College in the Strand.

The outstanding problem is whether the plan can be implemented in time to make the necessary savings. Schools that must merge will be combining courses, departments and administrations and reorganizing accommodation. St Thomas's and Guy's hospitals, however, will retain their separate schools but merge administration.

The merger of the London and St Bartholomew's schools will take longer. The promise by the University Grants Committee to finance new accommodation for the joint school at Queen Mary College, a multi-faculty institution, still stands. But the grants committee is now looking for existing buildings which may become available when Queen Mary College has itself trimmed its operations to match its reduced grant.

The reorganized medical schools will be expected to maintain their aggregate student population. But the reorganization alone will not by itself yield the necessary savings. So student:staff ratios are to decline from the present 1:7.4 in pre-clinical studies to 1:10 by 1983–84 and from 1:6 in clinical studies to 1:7.

Judy Redfearn

Polish sciences

Nothing to read

Poland's endemic hard-currency problems are posing a major threat to Polish science. The purchase of Western scientific journals is virtually impossible, and according to the Warsaw daily *Zycie Warszawy*, last year no subscriptions were paid to foreign suppliers at all, so that by October the country had run up a debt of about US$8 million for journals supplied against invoices.

This year, some $5 million have been allotted for the purchase of journals, but this is only half the quota for 1980 and less than 20 per cent of the 1978 level. Polish scientists have had to resort to various stratagems to keep up with their reading. Photocopies of *Current Contents* were displayed on the noticeboards of institutes and universities and scientists requested a photocopy of articles they required from the Academy of Science or else through interlibrary loan. Martial law, however, meant that very tight security controls were imposed on photocopiers, lest they be used for the production of protest leaflets.

At the same time, their other main source of new publications — offprints and duplicate copies of journals requested from foreign colleagues — has been considerably reduced. Under martial law, the mails are considerably delayed by the censorship, scientific visits to and from Poland have been significantly curtailed, and many Westerners are apparently disinclined to send material to their Polish colleagues for fear that journals arriving from abroad might attract the attention of the security authorities.

So far, little has been done abroad to relieve the situation. The British Council has allotted £25,000 for the purchase of journals for Polish academic institutions; in the United States, however, the view is by no means unanimous that aid would be proper until martial law comes to an end.

But to Polish scientists the situation is a matter of intellectual survival. Letters to Western colleagues emphasize that without an emergency supply of journals to tide them over the current crisis, Polish scientists will rapidly fall behind the world scientific community, and catching up, when dollars become available again, will be virtually impossible.

Vera Rich

0028-0836/82/210259-01$01.00

CORRESPONDENCE

The DNA disease

SIR — This is an account of a disease which has recently attained epidemic proportions. In February 1975 at a meeting held at Asilomar Conference Centre in California, many scientists and the world at large first learned of recent developments which came to be known as the recombinant DNA molecule technique, whereby genetic information initially in one organism could be translocated into another where, with reasonable luck, it might prosper and ultimately find expression. This announcement was followed by a wave of anxiety among the knowledgeable. It was confidently forecast that among the new chimaeras which might thus be generated there would be some that were malevolent. New toxic gene products would be let loose in the world and organisms which had, for millennia, lived in friendly symbiosis would be rudely ejected from their econiches by novel and hostile forms.

The assembled scientists at Asilomar therefore agreed to conduct their experiments under severe constraints of biological and physical containment. The intent of these regulations, drafted as "guidelines" by the Recombinant DNA Molecule Program Advisory Committee of the National Institutes of Health, was to minimize the likelihood of any of the foreseeable catastrophes.

It is now quite generally agreed that the anxieties which were initially generated were without much substance. Curiously, however, the one serious damage to our ecosystem which has now been observed to occur as a result of application of the recombinant DNA molecule technique was totally unanticipated in 1975.

The problem has arisen as a result of the discovery that some of the gene products derived from recombinant DNA experiments commanded a high market value. The gene which was enhanced was that for venality and the unanticipated gene product was money — sometimes in the form of shares in an entrepreneurial company — sometimes in the form of stock options. This gene product to which many investigators had not previously been exposed might well have been contained had the experiments been conducted under appropriate physical containment such as that in a bank vault. However, the new disease proved to be quite contagious, spreading rapidly even to the administration of the institutions in which these scientists were housed. With great eagerness, everyone wanted a piece of the action.

Among the symptoms of the disease most prominent has been a deterioration in communication, both verbal and written, on the part of those affected. They simply will not tell their friends and neighbours what they are doing or what they intend to do. Both teacher and student have schooled themselves to be close-lipped lest the competitor discover the name of the new product or the means to its production.

Concurrently, a change was noted in the dissertation problems assigned to graduate students. Whereas these formerly always contained an element of new *knowledge*, now the stress is more towards a new *product* and preferably a marketable one. The student may not be assured of what would formerly be termed an excellent graduate education, but he is more or less guaranteed a well-paid job at the end of the road. In some instances, if the entrepreneurial company is remote from the campus at which the professor is nominally employed, the academic chair which he is expected to occupy may be vacant much of the time. Science gives way to technology, research to development and publication to patent. Lip service, of course, is still paid to basic research, but clearly the product is the thing. For what is an entrepreneurial company without a product?

The prognosis for this epidemic is, at present, unfavourable. So far no treatment for the disease has been forthcoming; and whereas measures to prevent its further spread can be imagined, no one seems to have the strength or the authority to impose the rigid quarantines that would be necessary. Readers of the *Arabian Nights* may recall what happened when, inadvertently, the genie was allowed to escape from the bottle. Having allowed the venality gene to escape, we cannot foresee how the matter will ever be put right again.

DEWITT STETTEN JR

Department of Health & Human Services, National Institutes of Health, Bethesda, Maryland, USA

Turkish rights

SIR — As a foreigner who has studied and taught in Turkey for about eight years, I found your version of Dr Ögelman's case lacking rationalism and credibility (*Nature* 25 February, p.638; 18 March, p.186).

I was a member of the faculty of the Middle East Technical University at Gaziantep during the academic years when the internal political strife and violence on the Turkish campuses and in the streets was at its peak. It is no secret that some of the faculty at our campus, and indeed at every other campus, were openly involved in supporting the extreme right or left. As a matter of fact, during those days it was almost impossible and certainly very dangerous not to take sides, which meant inviting the wrath and enmity of both sides. At our university, where the leftist students and the faculty had the upper hand, anyone trying to be neutral and reasonable was branded as a fascist, which made the foreign faculty members, like myself and other North American and West Europeans, very vulnerable and insecure. My office and classroom were heavily pelted with stones for just beginning a lecture on the date when some leftist student had been jailed or hanged in the past. There were daily incidents of bombing and stabbing inside and outside university campuses and everyone lived in an atmosphere of terror.

Active collusion of a part of the faculty with the criminal elements of both the right and the left frustrated the efforts of the university administration to bring about normalization. The police, who could not enter the university premises without prior authorization from the university, usually waited outside as passive and helpless observers while mobs of unruly students of a majority group kept the other students outside the university grounds.

I am neither familiar with the facts of Dr Ögelman's case nor is it my intention to defend the policies and/or practices of the Turkish government. It is, however, a known fact that Cukorova University, which was only a couple of hours drive from us, was the scene of repeated violence and had to be shut down frequently. You state that "Dr Ögelman has been prosecuted not because she is a physicist but because of her part in an organization intended to secure rights for women in a society in which these have recently been denied".

I am afraid that the basis of your above statement as well as your advocacy for the cause of freedom for the Turkish scientific community need to be re-examined objectively and impartially. First of all, the rights of the Turkish women are solidly entrenched in the constitution. I have made several inquiries here and no one seems to know what rights the Turkish women have been denied recently and how Dr Ögelman's organization is planning to restore those rights. In this connection, it would have been helpful if you had included the name of that organization also. From my own experience and knowledge as a resident of Turkey for almost eight years, I have no hesitation in saying that the Turkish women enjoy more rights and have far greater opportunities for advancement than the women of North America.

As a scientist living and working in a free society, I appreciate and support your efforts in publicizing the plight of fellow scientists under oppression, but such incidents must be presented along with credible evidence which, in my opinion, has not been done in Dr Ögelman's case.

M.H. SADAR

Middle East Technical University, Gaziantep, Turkey

SIR — Your article "Turkish illiberality" (*Nature* 18 March, p.186) is a completely erroneous interpretation of the liberality in Turkey today. Being the co-author of the research paper published in the same issue of your journal (p.231), I must protest strongly at your misinterpretation of the situation of people with limited freedom in my country. Above all, the fact that you equate the Turkish martial law authorities with the Polish and Argentinian martial law authorities has driven me to think that your long established, serious and scientifically respected journal is taking advantage of its position by resorting to cheap politics.

As you should know, Turkish women gained their rights in 1923 under the great leader Atatürk. The women's organization mentioned in your article has attempted to give a misleading impression of the Turkish community, which consists of wholly liberated women. Scientists are human too; they can commit wrongs and rights, and can be condemned or set free.

The martial law authorities in Turkey have not limited liberality and democracy, but on the contrary have created an atmosphere of pleasantness and friendship between people who were enemies before the present regime took power. Above all, their existence has eliminated anarchy and its fatal consequences.

SELIM KAPUR

Adana, Turkey

SCIENCE IN WEST GERMANY

Discovery and disappointment

THIS year is a turning point in the development of science in West Germany since the Second World War. For one thing, this will be the first year in which government support for scientific research will not increase in real terms. For most agencies, the shortfall is not serious. In deutschemarks, indeed, budgets will be bigger, perhaps by four per cent. If inflation turns out to be six per cent, the real decrease will be a mere two per cent. But inflation could be greater than that, while, as people elsewhere know, official retail price indices are not an accurate guide to the cost of scientific work.

The budget cuts, although unimportant in themselves, are nevertheless symbolic of questions that have arisen in many minds. In the past three decades of heady prosperity, both the federal and regional governments have backed research generously, and for two reasons: the restoration of German scholarship is as necessary as the restoration of the economy and then, of course, science is an important source of innovation and thus a kind of guarantor that the economic miracle will not fade. So if even the West German economy can be driven into recession, may it not have been a miscalculation to suppose that investment in research would keep economic growth alive? The logic is not very sound, but the question, once asked, cannot easily be suppressed.

To politicians, the question takes the form "Are we getting value for all that money?" Their officials are responding with an energetic campaign to assist the process of "technology transfer". A few spectacular successes could make future budgets safe. But the question echoes another which the scientific community in West Germany seems also to be asking: How well are we doing, anyway? After all these years of effort, why is the United States the place where most of the interesting discoveries are made? Is something wrong?

The most surprising impression left by talks with a variety of scientists and officials is the tentativeness of science in West Germany. People seem to spend an alarming amount of time looking over their shoulders or trying to devise criteria for telling how well they are doing. (Nobel prizes per head per year? Mentions in *Science Citation Index*?) In reality, such lack of confidence is entirely inappropriate. West Germany is not merely crammed with scientific enterprises of various kinds, many of which are internationally known, but the sheer scale of the West German enterprise is breathtaking. Yet even if the state of mind is paradoxical, the reasons why it has arisen may be of absorbing interest.

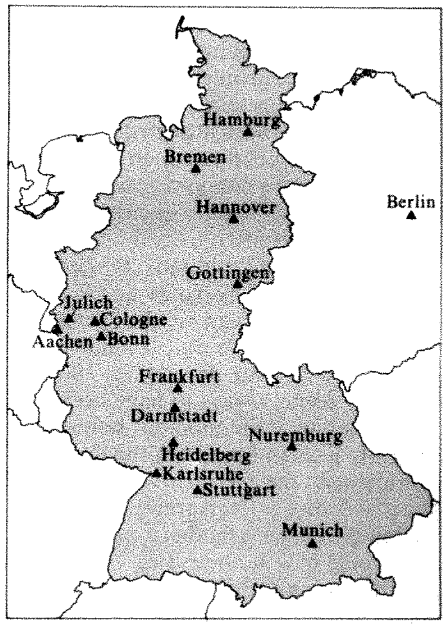
The brief survey of the state of science in West Germany that follows must not be mistaken for a catalogue. It is based on talks with officials of the federal government in Bonn and a string of visits to universities, laboratories and institutions at scattered centres as well as on the lavish use of the telephone. One objective is to investigate an interesting phenomenon in a manner that may interest readers. Another is the hope that it may be easier in future for *Nature* to do justice to an important part of the international scientific community.

As will be seen, much of what follows is concerned with the mechanisms by which public funds are channelled into research in West Germany. For all kinds of reasons but not least the interaction between the federal and the *Länder* governments and the federal government's constitutional compulsion to delegate decision-making to virtually autonomous bodies such as the Max-Planck-Gesellschaft, these mechanisms are more interesting than almost anywhere else. One consequence is that part of the scientific establishment has been able to capture part of the government's machine, and to use it itself.

Especially in these circumstances, why should there be such a general air of self-conscious uncertainty? (One explanation is that the responsibility is almost too great, but few agree with that.) The first thing to say is that these doubts crop up only in places where the objective is basic research, and where people *might* hope to win Nobel prizes even with the restricted definition of science included in Nobel's will. The people in the engineering school at Aachen know they are doing a unique job superbly, and have no recognizable self-doubt. (Chasteningly, one of them also asked "What is this journal *Nature*?")

Nomenclature

THE Federal Republic of Germany (*Bundesrepublik Deutschland*) is referred to throughout as West Germany so as to avoid the use of some set of initials (for example, FRG) that does not translate into German. The German mark (abbreviated as DM) is the unit of currency except when inappropriate. DM1 = £0.24 = $0.43 approximately. People's formal German titles are abbreviated by the omission of references to first degrees (for example, *Dipl.-Ing.*) and by the use of one or other (but not both) of the titles Professor and Dr. Place names are given in the English form when the German names cannot easily be pronounced by English speakers (for example, Munich, Cologne) but Hannover is spelled with two ens, as it should be.



So does the self-doubt have something to do with West Germany? That would not be a surprise. Although well over half of those now living in West Germany were born after the end of the Second World War, the economic miracle has tended to obscure the plain truth that West Germany as it now is was cobbled together by the three Western occupying powers in the late 1940s, became a fully sovereign state only in 1955 and gave up all hope of unification with its other half only in the late 1960s. And there remains West Berlin as a reminder to all who care to notice that the problem to the East was never resolved as intended at the Potsdam conference in 1945. As it is, three hours' driving (West German style) will take you from the Dutch border at the Ruhr to the eastern border near Braunschweig. From Göttingen, further south, you can walk to the border and back within the day. And if such circumstances fail to remind the young that West Germany is a living piece of suspended history, like a fly in amber, there are always people ready to remind them. One young woman in a Hannover restaurant asked whey she had *still* met so much rudeness on a recent visit to Britain. Who knows how to quantify the effects of this on people's confidence at the bench?

More particular versions of this question inevitably arise. What for example were the consequences of the exodus from German universities in the 1930s? And of the Catch-22 trick that in the 1930s encouraged public servants such as university teachers to join the Nazi Party and then, in 1945, denied public service to former Nazis? The result was that the post-war cohorts of undergraduates, now of an age to be laboratory directors, were mostly taught by people who had been unwillingly conscripted into military service. Again the importance of this effect cannot be quantified; the pages that follow will not

0028-0836/82/210261-02$01.00

attempt to answer the implied question.

One startling comment on this conundrum, from an older scientist, is nevertheless worth recording. The exodus should not have mattered, the argument goes. For the people who left during Hitler's time, Einstein and the like, were people who had been attracted to pre-war Germany from countries such as Switzerland, so that their disappearance mostly to the United States should in eugenic terms be strictly irrelevant.

Historically, and culturally, the first half of this century may come to seem an irrelevancy. For can it be simply a coincidence that West Germany has finished up a latter-day replica of what Germany seems always to have been — a confederation of autonomous states, as in the Holy Roman Empire of the thirteenth century and the German Confederation that followed Napoleon's defeat? Certainly neither Hamburg nor Munich now would have it otherwise.

Other manifestations of historical continuity keep cropping up. The past decade's spate of social legislation? Merely an extrapolation of the Weimar Republic's attempt to codify the liberally-paternalistic practices of the mediaeval German princes and their successors, the joint-stock companies of the nineteenth century. Legalism, obsession with laws and their interpretation? But that's straight from the Romans, and anyway you cannot feel free until your freedom has been defined. The tendency towards ideology (purpose shaping action) that prevents those who hold to the concept of the interdependent society from breaking into each other's cars but which tempts a small minority along the Bader-Meinhof path? If Calvin's reformation (when Geneva was still part of Germany) caused such trouble, it could not now be seemly to accept as colleagues people whose motives are misguided, whatever their other virtues (see page 264). And scholarship? Even now, in 1982, few can take up that subject without referring to Wilhelm von Humboldt, the founder (in 1810) of the University of Berlin, who at the same time put out the then-liberal and now supposedly-effete message that good research begets good teaching. Evidently the Third Reich might never have happened. And the West German universities have splendidly played their part in fostering this sense of historical continuity.

But for how much longer? It is clear that opinions in West Germany are deeply divided about the wisdom of what has been done to the universities in the past decade or so. Inevitably, the pessimists are more vociferous than the optimists, but they seem also to have the stronger case. So no apology is needed that much of what follows reflects the anxieties of those who are alarmed at the consequences of reform.

This survey has been compiled by John Maddox with assistance as indicated.

Will the miracle come back?

WHATEVER happened to the West German economic miracle, the phenomenon by which in the 1950s Chancellor Konrad Adenauer and his economics minister, Ludwig Erhard, arranged that West Germany should become the most prosperous of all the states of Western Europe? The most comforting answer is that the economic miracle is still there, just waiting for the long recession to end.

This is probably half true. Economic growth and large trade surpluses persisted throughout the 1970s, while other Western economies were already faltering. Only in the past two years has the growth stopped. Unemployment (once negative in that West Germany in the early 1970s relied on an estimated 2.3 million "guest-workers" for its labour force) had begun to rise in 1973, and is now heading for 2 million. And inflation, somewhat exaggerated by the West German retail price index at an average of 6 per cent a year in the 1970s, has now begun to increase. At least part of the trouble is that West Germany has at long last been affected by the stagnation of international trade in the manufactured goods which constitute some 47 per cent of the country's export trade.

Nobody, however, believes this to be the whole story. For one thing, there is the cost of oil and natural gas, almost all of which has to be imported. Although conservation measures have had a substantial effect on petroleum imports (down from 98 million tonnes a year in 1974 to just over 90 million tonnes a year now), the cost of oil and gas is between a sixth and a fifth of West Germany's import bill. This charge on the economy is not going to go away.

People in West Germany are more alarmed, however, by what seems a sinister threat to future prosperity — the challenge from Japan. The days have long since gone when Volkswagen's beetle swept the world. Now the beetle's successors have to compete in the export markets with cheaper but still well-engineered products from the Far East. The same is also true in electronics. And for other reasons, the West German chemical industry — the other foundation of the post-war export trade — has faltered, financially shackled by expensive investments conceived of when economic optimism was the order of the day. So everybody is convinced that there must now be a new period of creative innovation in West German industry. "Technology transfer" is the slogan. Science may be one of the beneficiaries.

Much less often, people also worry about the possible consequences of the past

Briefing

WEST GERMANY is the quintessential federal state. Its existence springs from the decision of the three Western occupying powers (France, the United Kingdom and the United States) in 1947 that a measure of self-government should be permitted in the west of the country pending agreement on German unification. The "Basic Law" (*Grundgesetz*) adopted at Frankfurt in 1949, for practical purposes the constitution of West Germay, reflects the insistence of the occupying states that political power should as far as possible rest with the regional (*Land*, pl. *Länder*) governments. (West Berlin is an eleventh *Land*, but has a special status.)

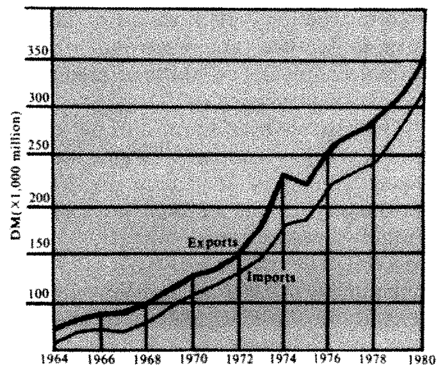
The basic law provides a complicated system of checks and balances by which the powers of the federal government are constrained. The power to take on tasks (such as research) not specified by the basic law requires a formal treaty between the federal and the *Länder* governments. The second chamber (*Bundesrat*) of the federal parliament consists solely of representatives appointed by the *Länder*. The federal constitutional court (*Bundesverfassungsgericht*), based at Karlsruhe, interprets the constitution on demand, and can overturn legislation deemed unconstitutional while also considering, for example, appeals by students against examination results.

The partition in 1945 of Germany and its consequences (east-west migration) have left West Germany with just over two-thirds of the total area (excluding the parts of Germany transferred to Poland in 1945) but four-fifths of the population (now 61 million).

The post-war history of West Germany has been shaped as much by the economic miracle of the early 1950s as by the almost continuous wrangle about East-West relations. Starting from an admittedly low base, for the best part of a decade the West German economy grew faster than any other in Western Europe. Now the Gross National Product is close on US$10,000 per head of population, close to that of the United States.

Since 1969, the federal government has been formed by a coalition of the Social Democrats (SPD) and the Free Democrats (or liberals), with Willy Brandt and Helmut Schmidt as successive chancellors. Now, however, there are signs that the coalition's hold on the federal government (and the cohesion thereof) is weakening. The Chancellor has had trouble within his own social democratic party on nuclear policy (civil as well as military), while the coalition has begun to lose control of some *Länder* governments. The election in Hesse (capital Frankfurt) in September is likely to be crucial to the coalition's future (and to its capacity to govern effectively between then and the *Bundestag* election still two years away).

0028-0836/82/210262-04$01.00



The increase (in actual values) of West Germany's exports and imports. The major exports are all types of machinery, motor vehicles, electrical engineering and chemical products. Major imports are food, drink, tobacco, petroleum and natural gas.

decade and a half of social reform introduced by the Social Democratic Party (SDP). Certainly the past few years have seen a substantial increase in social transfer payments by the federal and *Länder* governments for such things as pensions, student support and unemployment assistance (and the high cost of these statutory payments when tax revenues have been reduced by the recession is the immediate cause of the economies now being enacted throughout public administration). But there is no sign that the SPD government's reforms have in any way impaired the efficacy of the market economy which has been the engine of economic success in the past three decades. Thus the statutory extension in 1976 of the rights of workers to participate in management decisions (at a time when even a British Labour government failed to make headway towards similar objectives) is said even by Christian Democrats to be but a continuation of a tradition going back, in West Germany, to the 1920s.

But may not people's attitudes have changed? The economic miracle was a heroic time for West Germany. People rose to the challenge. Now, some fear, affluence has made the younger generation work-shy, from which point of view the activities of the "greens" (politically active environmentalists) is akin to sabotage of the West German dream. Such gloomy after-dinner musings are probably mere speculation. That they are offered at all as serious contributions to the question of what the future holds is a measure of the extent to which the economic miracle lies in the past. □

Universities reformed to death

By spitting out a peppermint in a dental surgery, does a dental student demonstrate such an ignorance of simple hygiene as to be unqualified for his profession? Last year, after an angry scene with such a student, a member of the dental faculty at Munster took this view, and said that a final examination should be retaken. But the local appeals court has now decided in favour of the student.

This solemn affair illustrates one of the common complaints of academics in West Germany. Legal interpretations of the public duty of academics (who are public servants) take precedence over research and teaching. But of the thousands of suits brought against the universities each year, most come from would-be students seeking to prove that they have been wrongly denied a place.

So what has become of the "Humboldt university", that community of scholars willing to share their learning with anybody who chose to take part? The embittered say that the classical German university has been killed off by the sequence of reforms to which the West German system of higher education has been subjected in the past fifteen years. The reformers agree that the system is still in turmoil.

The underlying trouble is that West Germany has sought to reconcile several irreconcilables — the principle of open access to any university in the country, the doctrine that all universities are equal, the practice that universities are run by the ministries of culture in the *Länder* in which they happen to be sited and the phenomenal increase in the demand for higher education in the past twenty years. Add the West German conviction that the law is the law is the law, and the plain fact that the reforms, consisting as they do of formal agreements between the federal and the *Länder* governments, are public legal documents, and you have a recipe for muddle, even chaos.

Historically, this mess was preordained, After the twelve years of incivility that ended in 1945, what can have been more natural than that the few surviving academics and the army of their would-be successors should have set out to recreate the Humboldt university, with all its blinding virtues and glaring faults?

The principle is admirable. The university is a community of scholars supported by the state (*Land*) as if it were an opera company. The scholars or professors (*Ordinarien,* sing. *Ordinarius*) *are* the university but each year they elect one of themselves (the *Rektor*) to be their head. As public servants, their duties are spelled out (for example, to lecture eight times a week) but they negotiate their salaries with the *Land*'s appointed bagman in the university, the Chancellor. He and not the *Rektor* also decides each professor's allocation of assistants, research expenses, janitors and capital costs. Students do not explicitly enter these calculations. The bright or pushy become unpaid laboratory assistants ("learning through research"). The others move on elsewhere, looking for something more congenial.

This system, which made possible the flowering of German scholarship at the end of the nineteenth century, is obviously open to abuse. Well-favoured professors may acquire delusions of divinity, and may tyrannize their assistants. And since even among scholars, rivalry and even jealousy may from time to time occur, professors in related fields and their respective entourages, organized into separate institutes, may occasionally behave like warring empires.

Great men and good universities usually managed to avoid such dangers. But the universities recreated after the Second World War were faced with a rapid growth in the numbers qualified to be students. By 1960, it was plain that the then-existing universities could not be expanded indefinitely, and that new institutions would be necessary. With Nordrhein-Westfalen in the lead, and with the promise of financial assistance from the federal government, the *Länder* set out to create new universities more or less in the image of the old. But the pace of construction was insufficient, and the replication of universities in the existing pattern could not meet the needs of society, let alone of students.

So what was to be done? By the mid-1960s there was a fierce argument between those represented by Professor Ralf Dahrendorf (then at Cologne, now director of the London School of Economics) who argued that there should be institutions equipped to provide vocational tertiary education (and shorter and more structured courses all over) and those who sought a more radical change in the structure of the university while preserving the cherished notions that all universities are in principle equal and that qualified students may study where they like.

The argument, unfortunately, did little to help the universities overwhelmed by students, some of which began to operate a *numerus clausus* rule — students should be turned away if there were already too many of them in the university or some part of it. This offence against the doctrine of free access was held to be an infringement of people's constitutional right to follow the profession of their choice. The operation of a quota system now, in fields such as medicine and veterinary medicine, is one of the chief sources of litigation by would-be students.

The second wave of expansion in higher education that began in 1970 appears to have given something to both parties in the earlier argument for reform; technical colleges of various kinds, traditionally means by which the graduates of non-*Gymnasium* high schools acquired a vocational education, were enlarged, upgraded and called *Fachhochschulen.* And the new universities established after 1970 (often built around pre-existing technical or teachers' colleges) were "comprehensive" in that they offered courses of frequently vocational study,

A conservative opinion on reform

THE universities are intrinsically conservative, so how can "reform" imposed from outside fail to damage them? This was the gloomy lament of *Rektor* of the University of Munich, Professor Nikolaus Lobkowicz, on the day last month when he had just handed in his resignation to the Bavarian government after being voted out of office.

A touch of bitterness would in the circumstances have been forgivable; that university rectors should be elected every four years is laid down in the federal reform law for higher education. But that suspicion is too dark. Professor Lobkowicz is a political scientist born in Czechoslovakia who has spent most of his working academic life in the United States. His description of himself as a true conservative is relevant.

The University of Munich is the biggest in West Germany, with 45,000 students or thereabouts. Bavaria is the stronghold of the Christian Democratic Union and, Lobkowicz thinks, the *Land* made a tactical mistake in 1969 by trying to preempt the federal law then being hatched by making one of its own. Bavaria overlooked its own long-standing tendency towards "preposterously perfect" legislation.

Lobkowicz — eloquent on reform

Legalism has run riot. In one recent year, the university awarded 1,100 doctoral degrees but had to fight 1,300 court cases, most of them brought by students contesting examination results. Occasionally, the university takes the government of Bavaria to the Constitutional Court in Karlsruhe.

The bureaucracy, the *Rektor* says, is palpably a pain; the damage done to academic standards cannot be as easily assessed. But surely the old system, in which holders of a university chair (*Ordinarien* in the old nomenclature) were treated as gods in their departments, was capable of corruption? So what? is the surprising answer. "Who says that corruption excludes excellence?"

The most serious of the present dangers in Professor Lobkowicz's opinion is that the new system for appointing people as professors of one of the Western world's great universities will ensure that they are "smooth" people, first of all electable and only secondarily capable of scholarship. The drill now is that the senate instructs the *Rektor* to submit a list of three nominees for each professorial appointment to the *Land* government, in order of the "soviet's" preference. (Students and other ranks are represented on these committees, but professors have a majority.)

The minister knows that there will be a row if he fails to appoint the first name on the list, but apparently there have been occasions when a discreet telephone call has persuaded him to take the risk. One of Lobkowicz's own regrets is that he failed to persuade his fellow academics to appoint as an honorary professor a distinguished Marxist philosopher "with whose every word I disagree" but who might have been good for the university. Reform, says Lobkowicz, has made the world safe only for the safe.

The result, he says, is that both scholarship and the scholarly professions have been debased. The great teachers of the old days have been replaced by people on the hunt for material with which to embellish their *curricula vitae*. And the days have long since gone when it was reckoned better to be a university professor than a bishop. So university teaching has become less attractive to bright people, those left are reluctant to welcome colleagues who may be brighter than themselves — and the brightest students are in any case persuaded to opt for courses of study to which entry is restricted, medicine for example.

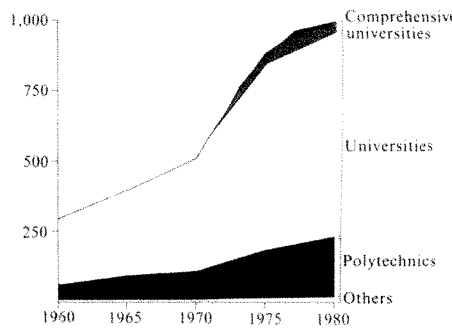
On the state of West German science, Lobkowicz is more puzzled than dismayed. Nobel prizes or the purchase of foreign patent rights are poor criteria of success, he says, while the prevalent academic convention to despise the popularization of scholarship has removed academic scientists from intelligent appraisal by their fellows. He notes the growing tension between university teachers and those working in Max Planck institutes, who have more time for research. He agrees that young people returning to West Germany from a successful spell in, say, the United States tend to lose their sparkle once back at home.

The explanation? The bureaucracy engendered by the reform. Yet Munich has been lucky in its minister of culture — it has dutifully "integrated" the *Pädagogische Hochschule* and only one department has been "ruined" in the process.

Why not, then, reform the reform? The outgoing *Rektor* is despondent even about that. For an amended law would be even more detailed than the present, allowing more scope for legalism. But Lobkowicz says that he differs from his colleagues elsewhere in this opinion and he insists that his is a conservative view. It is, however, eloquent and influential.

used instructional methods other than the formal lecture yet awarded degrees formally equivalent to those offered by the pre-existing universities.

The cap has been set on this pattern by the federal government's "frame law" on higher education, a general prescription for the ways in which *Länder* governments should organize their universities made constitutionally possible by the amendment in 1969 of the West German constitution, which permitted collaboration on "common tasks" between the federal and *Länder* governments on



Student registration in higher education in West Germany from 1960 to 1980 (in thousands).

matters previously reserved to the latter. One objective of this tortuous piece of legislation is the removal of the opportunities for abuse in the classical German university. Departments rather than professorial chairs (*Lehrstühle*) were decreed to be the "basic organizational units", *Rektors* were to be elected for four years rather than one, staff vacancies were to be advertised, and so on.

The crucial and controversial part of the frame law, however, is the declaration that

> . . . the different types of institution of higher education shall be brought together to form a new system of higher education. Institutions shall be extended or merged to become comprehensive universities or, while retaining their legal autonomy, shall be linked together by the establishment of joint bodies to form comprehensive universities. Where it is not, or not yet, possible to establish comprehensive universities, cooperation between institutions is to be assured.

The zeal with which these principles have been prosecuted varies from *Land* to *Land*. Although the deadline for decision specified in the act was 1979, the system is still in flux. Observers from outside must marvel at the confusion that results. Here is one list of what seems to be awry:

- A university cannot match student registration to its capacity to teach unless the *Land* government has formally declared a shortage of capacity, in which case only the Central Office for the Allocation of Study Places can decide which students shall be taken in. In practice, quotas apply principally in medicine and related courses. The law specifies that two-thirds of the available places should be allocated according to

grades in high-school leaving examinations, but the number of years spent waiting for a place should count in the allocation of the remainder.

- The time spent by universities and their academics in providing information about the way they spend their time (essential if a university is to defend itself against law suits brought by students who argue that they have been turned away when they could have been taken in) is an intolerable intrusion that makes the "effort reporting" required of grant-holders in the United States seem benign.
- While entry to restricted courses is to some extent determined by school-leaving grades, the recent reform of the *Abitur* system makes possible a degree of specialization at school, but there is no means of linking that with the course chosen at university; whence the complaint that school-students plump for easy options when their real aim is to enter a medical school.
- The attempt to modernize the content of university courses, and to ensure national uniformity in deference to the conviction that a degree is a degree (or that all universities are equal) and to ensure transferability from one place to another has the predictable effect of increasing the duration of students' courses, now more like six years than the minimum four. (Students' grants, a liability chiefly of the federal government since 1972, run out after five years.) Yet students' preferences for the older universities suggest that they do not regard the universities as equal, so the newer universities are, through no fault of their own, under-used.

The consequences of all this for the conditions of the research enterprise are only indirect. The mechanisms for financing research in universities are enlightened (see below), and almost entirely in the gift of the research community itself. But in the universities that remain popular with students, the teaching load has become intolerable (research during vacations only), and the nature of academic life has changed in such a way that people are likely to find the Max Planck institutes even more seductive than they already are. Doubts persist about the way in which students are being taught.

The past few years have, however, seen a more immediate problem. The federal and *Länder* governments, linked financially together by the taxation system, are being forced to live within their means. In many places there is a moratorium on filling academic places, and a 7.5 per cent reduction of all public posts is likely by 1986.

In Nordrhein-Westfalen, the spearhead of educational expansion in the 1960s, the problem is more acute. By being first in the movement to establish comprehensive universities, the *Land* has saddled itself with educational facilities still under-used, as well as with institutions such as the great medical school at Aachen (see page 267) whose future must be problematical. The result is that the word went out from Dusseldorf (the state capital) at the end of March that there must be reductions of the scale on which the traditional universities operate. In a form reminiscent of what has been happening in Britain in the past few years, the Minister of Culture has recommended that this or that department must be closed. The University of Bonn, now bursting at the seams with 35,000 students (including 5,500 law students) in buildings fit for half that number, has been told that it must lose part of its teacher training programme, and its institutes of egyptology and Scandinavian studies. Dusseldorf's immediate target is to save 400 jobs at Bonn. Nobody knows how many of those posts are intended to finish up at places such as Aachen. Indeed, it is still possible that the overcrowded universities will be able successfully to resist what Dusseldorf has in mind. If the issue is settled before the mid-1980s, when the falling school rolls will have worked their way through to higher education, giving some relief to the overcrowded universities, everybody will count himself lucky. □

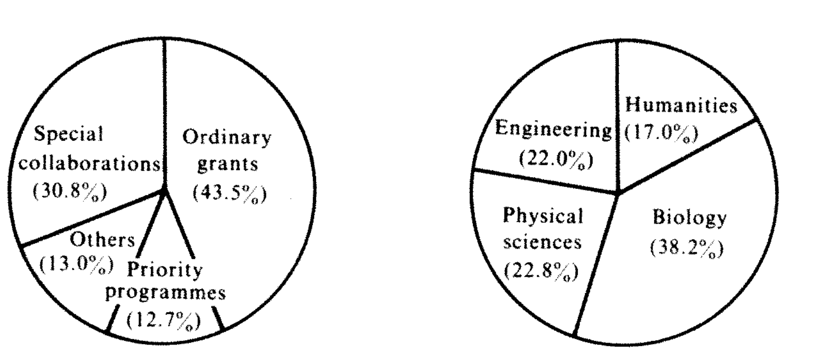
Where to get research grants

PLURALITY is most academics' defence against the peer review system, but not in West Germany. For there is just a single agency for supporting research in higher educational institutions — Deutsche Forschungsgemeinschaft (German Research Society) or DFG. Curiously, however, this single organization is not the monolithic enterprise that might be feared; grant applicants speak well of what it does, even if university administrators complain of having to sign several identical pieces of paper for each grant application.

Part of the explanation for the respect accorded to DFG is historical. There has been a device of some kind for helping academics to carry out research since 1920, when Max Planck and Fritz Haber persuaded the then *Reich* to put money into such an enterprise. After the Second World War, DFG was formed by a merger of two similarly intentioned organizations as the result of a treaty between the federal and the *Länder* governments.

The legal constitution of DFG also helps. It is not a government agency but the strict equivalent of a "not-for-profit" organization in the United States or an operating charity (a company limited by guarantee) in the United Kingdom. And those responsible for its operations are not people but a motley collection of legally autonomous academic enterprises — many (but not all) of the universities, some (but only a few) of the national research centres and a sparse sampling of learned and scientific societies. The result is that even though the funds for DFG's operations come from the federal government and the *Länder* governments, which are all represented on various high-level committees, these same committees provide such a complicated system of checks and balances that the potential beneficiaries of DFG have a real sense of making policy, particular and general.

At the same time, DFG has many of the functions of what are elsewhere known as national academies. Thus it pays the West German contribution to bodies such as the International Council of Scientific Unions, looks after the exchange of scientists with other countries in Europe and elsewhere, worries about scientific libraries (and such questions as computerized information services), carries out studies at the behest of government and runs the "Heisenberg programme" — West Germany's device for helping people who would in other times walk straight into academic jobs but who are now, in a period of relative stagnation, kept waiting on the sidelines. It differs from a learned academy, however, in the scale of its operations — the budget is close on DM900 million — and in not electing people to any fellowship. Naturally, German culture being what it is, DFG looks after the whole of scholarship, both Protestant and Catholic, theology included.



Left: Expenditure by DFG (in 1980) on the various forms of grant support. "Others" includes central facilities, administration and so on. Right: Distribution of DFG grants (in 1980) between fields of research. "Biology" includes medicine.

0028-0836/82/210265-02$01.00

Another Catch-22?

THE West German government, a great believer in economically counter-cyclical devices, was among the first to introduce (in 1978) a scheme for catering for younger scientists for whom there would in normal times have been empty academic posts, but who in present circumstances are likely to be forced out of research. How well is this Heisenberg programme working?

By the beginning of 1982, 246 people had been awarded fellowships, which can run for five years. The scheme was devised by the West German Rektors Conference, is supported financially by the federal and *Länder* governments, and administered by the DFG.

Those elegible for support must be qualified (by means of the *Habilitation* process), for appointment as a tenured university professor. One striking proof for this necessity is the frequency with which visiting cards refer to this qualification, unknown outside West Germany.

However, the DFG is apparently not able to spend all the funds at its disposal for this purpose — enough for 150 new fellowships each year. The problem is that selection committees are anxious that those awarded Heisenberg fellowships should be people who could, without disgrace to themselves or anybody else, move into a university post should one become vacant. But those are also the people who are likely to be snapped up by Max Planck institutes, industry and, even in these straightened times, the universities.

What seemed, only a few years ago, to be an imaginative solution to a widely recognized problem has thus turned out to be another Catch-22.

As a grant-making organization, DFG does business in a variety of ways. At one end of the scale is the standard procedure by which academics can apply for research funds to employ research assistants, to travel abroad, to be freed from teaching and research for a year and even to prepare (over a period of no more than two years) for qualification as a full university professor (*Habilitation*). DFG spends close on half its budget on these activities, and makes more than 5,000 grants a year with an average cost of about DM80,000 each. Half of the cost is met by contributions from the *Länder* and federal governments.

As in Britain, grants of this kind are made on the assumption that university applicants already have access to the essential tools of research. Hitherto, the *Länder* governments appear to have been generous in this respect, partly from regional pride, partly because well-equipped universities are better able to compete for centrally administered funds. But DFG is alarmed at the signs that have become apparent in the past two years that many applicants are no longer able to satisfy their sponsors on this point. With misgivings, DFG is now paying for some equipment that it would previously have expected the university to provide; the fear is that too much of this will erode its capacity for making grants.

Larger projects are financed by different mechanisms. A group of researchers at a university can apply to be financed as a research group, and there is a "priorities programme" that enables groups of academics anywhere in West Germany to collaborate on a research topic considered by DFG to require special attention.

West German participation in the Deep Sea Drilling Project has been in part made possible by a priorities programme grant of more than DM20 million to a research plan coordinated by Dr Helmut Beiersdorf from the Federal Geophysics and Natural Resources Laboratory at Hannover, which involves some 36 investigators at 14 universities (most of whom have used DFG funds to buy equipment). More recently, nationwide collaborations on this pattern have been set up in gene technology and the mechanical engineering of miniaturization technology.

The most distinctive of DFG's mechanisms for supporting academic research is, however, the scheme for providing long-term grants to university departments prepared to concentrate their research in some specified field. This *Sondersforschungbereich* programme supports projects as different as the University of Bonn's continuing interest in radioastronomy (see page 274) and a variety of clinical projects scattered through West Germany. Since the beginning of the programme in 1968, about 180 of these special collaborative projects have found their way onto DFG's books. The scale of support appears to range from DM1 million a year or less up to DM12 million a year (for the marine construction project at Hannover), with the average cost at about DM1.5 million. Unlike DFG's other grant-making activities, the costs of this programme are not shared equally by the federal and *Länder* governments but, rather, are divided in the ratio 3:1.

While it seems that DFG has frequently brought such projects to an end, there has been a tendency over the years for the number of current projects to increase. But now DFG is constrained by a recommendation of the Wissenschaftsrat that these projects should not normally last for more than 12 or 15 years. With 40 current projects going back to 1968, the next few years will see rapid change. This is part of the reason why the theoretical mathematics *Sonderforschungbereich* at Bonn is being converted into a Max Planck institute. Elsewhere, some discreet renaming of research projects may be necessary.

At the Max Planck institutes, DFG operates a formal and stringent system of invigilation for deciding whether support for one of these special programmes should continue. Every three years or so, a team of people in the field spends two days interviewing the members of each collaboration, raising awkward questions about the inevitable gap between aspirations and achievement. The invigilators pride themselves on being tough.

Broadly speaking, DFG seems confident that, despite this year's interruption, its previously steady growth will soon be resumed. Between 1981 and 1982, the budget grew by just six per cent, the equivalent of inflation, but the prospect is that next year's budget will be only four per cent greater than now — an increase less than the probable inflation rate. The general secretary of DFG, Dr Carl Heinz Schiel, expects that DFG's budget will continue to grow at least until around 1990, when demographic change may imply that the university system begins to shrink.

As an essential component of what would elsewhere be called the dual-support system for the universities, perhaps the most immediate threat is that, as in Britain, regular support for the universities may be so cut back by the *Länder* governments that DFG's grant-making capacity will be undermined. Inflexibility is another difficulty. Because of the different degrees of the involvement of the *Länder* governments in DFG's general and special programmes, the transfer of funds from one part of the budget to another is constrained. Moreover, because new *Sondersforschungbereich* projects have to be approved in advance by the Wissenschaftsrat, speed cannot be assured nor political interference ruled out.

What seems most conspicuously to be missing from the pattern of DFG's activities is some mechanism for supporting substantial programmes of research by the younger members of academic faculties. The Heisenberg programme (see box) is intended to keep some such people in an academic orbit when they might otherwise be frozen out, but the needs of young and potentially daring researchers seem to fall between the two stools of modest individual support as part of the standard grant mechanism and the much larger programme grants in which large groups of people, many of them senior scientists, are simultaneously involved.

The consequence may be the explanation given by one academic molecular biologist, a member of one of DFG's research groups, for his fatalistic expectation that the next dramatic development in his field would be made by some American group. But why not you? "Well," he said, "our hierarchy is more rigid than theirs, we are less mobile and our communications with each other are less good." Add to that the teaching load, which keeps senior academics away from the bench except during vacations, and you have a recipe for being second. □

A little knowledge

If *wissen* means "to know", then *Wissenschaft* means "knowledge", and since German culture makes no distinction between different branches of knowledge, *Wissenschaft* accordingly includes "science" as one of its English meanings. But this English legend does not correspond with present usage.

A further complication is the use of the prefix *Natur* (meaning "nature") as a qualifier, for *Naturwissenschaft* means "physical science", as distinct from, say, the biological sciences (as in *Biowissenschaften*), but *Naturforschung* appears to mean scientific research of any kind. Confusion is not common, but appears to have led to the common practice of translating *Wissenschaftsrat* as "science council".

That unique institution, made necessary by the federal character of the West German constitution and set up in 1957, has a general influence on the development of the university system and the research enterprise in West Germany. The *Wissenschaftsrat* is a kind of constitutional rubber stamp. Its general assembly consists of two constituencies — one appointed by the federal government and one representing each *Land* equally — with equal voting power. There are also outsiders from universities and business who do not vote. Technically, the council is advisory to the federal and *Länder* governments. Recommendations require a two-thirds majority, which means that the federal government can have its way if four *Länder* support it.

For the first decade and a half of its existence, the council was creatively influential; and even now it would be imprudent of the federal government or one of its agencies to think of doing something new without the council's agreement. Thus proposals that public money should be spent on a new laboratory, or that some new university should be established, are at some stage cleared with the council.

The agenda for discussion has shrunk since the constitutional amendment of 1969 that gave the federal government a role in higher education. Even so, it retains a life of its own. In the past two years, it has for example argued successfully that the *Deutsche Forschungsgemeinschaft* (DFG) should not make long-term programme grants for more than 12-15 years (which advice appears to have been taken seriously). The council was more recently influential in the decision that the new Alfred Wegener research centre for polar research should be located at Bremen rather than at Kiel.

The most accurate translation of the council's German name may therefore be "the council for the politics of scholarship".

Aachen's other cathedral

THE great *Klinikum* of Aachen stands in relation to the city at which Charlemagne was defeated as did the opera house to Sydney a decade ago: it is too soon to know whether it is one of the wonders of the modern world, a white elephant or both. But the now-estimated cost of DM 1,700 million of what may yet become the world's most gargantuan teaching hospital has already so cramped the budget of Nordrhein-Westfalen that the repercussions will be felt next October in all the *Land*'s universities from Bonn to Bochum.

Like Aachen's cathedral (probably the loveliest of them all), the *Klinikum* lies in a hollow but towards the west of the city, within sight of the Belgian and Dutch borders. Like all cathedrals, the *Klinikum* is huge, with a usable physical capacity of 130,000 m^3 and a gross volume (including air-conditioning plants and the like) of half as much again. If ever completed, it will house 10,000 people, including 1,500 patients, 2,500 medical students and up to 500 dental students. Like many mediaeval cathedral projects, the *Klinikum* has been plagued by subsidence — some parts of it have sunk by as much as 5 cm.

Like any cathedral, the *Klinikum* is the physical expression of somebody's vision — in this case, that of Mr Hans Wertz, once financial controller of the city of Aachen, who, when finance minister of Nordrhein-Westfalia in the early 1960s, appears to have wished on his native city and on the university (the Rheinisch-Westfälische Technische Hochschule) which the *Land* "owned" there, the promise of a building in which physicians could make their dreams come true, together with an accompanying blank cheque.

Although RWTH, as it likes still to be known (see page 275), now seeks to distance itself from the huge cost of the *Klinikum*, at the outset it was a willing accomplice.

This distinguished institution, best known for its contributions to the health and wealth of the West German machine-tool industry, acquired a medical faculty only in 1965, three years before the notion of the *Klinikum* was hatched. Even academics would have known that most other universities with medical schools could largely finance them from the payments made by the insurance companies with which West German workers are compulsorily insured.

During the past fifteen years of construction, the dream has faded in several ways.

- The hope that the Aachen *Klinikum* could become a truly European hospital, drawing patients from Belgium and the Netherlands as well as West Germany, has been frustrated because insurance companies' payments are no more transplantable now than in the 1960s (and because the Dutch have closed their border since terrorism became rife).
- The discovery that the infective organisms of legionnaires' disease flourish in the humidifiers of air-conditioning plants has made it necessary to modify the design of the large single plant.
- There was a successful prosecution in 1976 of a suit in the West German courts by complainants holding that work-people should not be required to spend more than four hours a day in rooms without windows. The result has been a three-year delay while the architects, Weber, Brand and Partners of Aachen, arranged that the internal partitions between the 6,600 rooms of the largest of all man-made caverns should be replaced by glass (which has, in turn, often been covered by venetian blinds).
- The main contractor for the *Klinikum*, the non-profit organization *Neue Heimat Städtebau*, set up as a cooperative of labour unions in the construction industry, has been involved in a generalized upheaval so widespread it may yet turn out to be the biggest socio-political scandal in the short history of West Germany. (The investigating commission has not yet reported.)

The result is that the *Klinikum*, surrounded by between 5 and 10 square kilometres of still-empty car park, remains unused. But the building is already being

Inside Aachen's great Klinikum *— but will the patients come?*

0028-0836/82/210267-02$01.00

used for teaching physiotherapists and it is hoped that 80 third-year dental students will arrive in October.

Academics on the Nordrhein-Westfalia payroll half snigger when asked "Why did you let this happen?" "Nothing to do with us", they say, "and anyway the *Klinikum* at Munster also cost a packet". But responsibility does not all lie in Dusseldorf. The university senate was willing enough to take over the city *Klinik*, offering to turn it into a medical school. The architects at Aachen were given not merely an opportunity to build a latter-day cathedral but a chance to try out a theory — the belief that construction and planning could proceed simultaneously.

Some of the immediate consequences are absurd. The new medical faculty is having to teach in the old *Klinik*, now hopelessly run down. And as the day approaches when the hospital may have real patients, perhaps in 1983, the university is alarmed that the *Klinikum*'s history (and appearance) will frighten off the patients from whose health insurance policies the operating cost must come.

The *Klinikum* is thus a monument to the economic optimism of the 1960s. Of the three medical centres then planned by the government of Nordrhein-Westfalia for its bemused and then fast-growing university system, that at Munster has been completed and that at Essen abandoned. And nobody knows whether West Germany needs all the physicians now being trained, some 5,000 a year. □

Where the power lies

IF there is a power centre in the federal government, with sway over the conduct of science, it must surely be the BMFT. It has a bigger budget than the education ministry, a more personable minister (Dr Andreas von Bülow) and a cleaner brief: spend money and make things happen. BMFT also finances the Max Planck Gesellschaft (see below), the most powerful but the most contentious basic research enterprise in West Germany. The trouble, unfortunately, is that not even the BMFT is as free as it likes to think.

Part of the problem is that the ministry is a tiny organization (and nine of that small number of posts were abolished when last autumn's draft budget was revised in February this year). Taxpayers paying taxes outside West Germany would be alarmed at the prospect of entrusting such small groups of people with such large sums of money. They would also be surprised to learn how civil (in the sense of being open) civil servants become when saddled with the responsibility.

Another part of the problem of telling where the power lies is how that power is distributed. The *Länder* governments have a lot of it. So too do the committees that have been given some kind of constitutional blessing by means of formal agreements between the federal and the *Länder* governments. Although the influence of the *Wissenschaftsrat* has declined since 1968, its formal recommendations must still be considered.

People also matter. West Germany being the place it is, a speech by some establishment figure, say the president of the DFG or of the Max Planck Society, will carry a great deal of weight. West Germany's preoccupation with the quality of research stems from a public speech by distinguished people who have left the public service and then made their misgivings public.

So science policv is determined in West Germany much as elsewhere, by accretion. The civil servants will defend the existing budget, but will leap at whatever opportunities for innovative expenditure are approved of by the councils of wise men.

Thus the BMFT is the sponsor of West German interests in space, energy, information technology, marine and polar research, while through its sponsorship of the national laboratories (see page 279) it is also directly responsible for much fundamental research, especially in particle and plasma physics. From one year to the next, it is possible to add or subtract relatively small amounts from these large programmes — this year, for example, an extra DM12 million for biotechnology (a 28 per cent increase) and an extra DM40 million for information technology (a 42 per cent increase on last year). More radical changes would be difficult. □

Centres of excellence

THE Max Planck Society (Max-Planck Gesellschaft zur Förderung der Wissenschaften) is internationally the best-known of all West German research organizations. *Everybody* knows *somebody* working at one of the 49 institutes scattered through the country. And that is one of the most frequent complaints about the society: if anything, the "Max Planck" is too successful, providing scientists with opportunities for research far better than those at the universities, for example.

But this is inevitable, given the constitution of the society . Historically, it is the post-war replacement for the Kaiser Wilhelm Gesellschaft, the organization established in 1911 whose Berlin institute provided Einstein with a base in the 1920s. The underlying objective is frankly elitist — to provide distinguished researchers with resources sufficient to support a group of younger colleagues, and to provide all concerned with the security and the equipment necessary for high-quality research.

On many occasions the Max Planck institutes have amply justified their founders' faith. The institute of biophysical chemistry at Göttingen for example, is where Professor Manfred Eigen's use of pulsed lasers pointed the way to novel techniques for following fast chemical reactions (and won a Nobel prize).

Flexibility is plain. The Göttingen institute has grown to the point at which there are nearly 400 people, many of them visitors from overseas, while the research programme has diversified enormously. This illustrates the Max-Planck contention that once an institute has been established, those in charge are within reason free to follow their changing interests.

Elsewhere, new institutes seem to spring up by mitosis. Thus the clutch of physics institutes at Garching, near Munich, can be traced back to the unique arrangement by which Professor Werner Heisenberg's Institute of Physics was financed in the years immediately after the war. By a succession of fissions, there are now for practical purposes five descendent institutes, one of them the huge plasma physics institute (see page 271) and the others concerned with physics, astrophysics, extra-terrestrial physics (see page 274) and quantum optics.

The Max-Planck Gesellschaft claims also to be vigilant in bringing institutes to an end when they have outlived their usefulness. Again (as with DFG) there is a system of formal invigilation every two or three years, with scientists from elsewhere included among the visiting party. But the list of casualties among institutes is not long — five have disappeared since 1948, while the thriving coal research institute at Mülheim (in the Ruhr) has become virtually self-financing by means of contract research.

While some Max Planck institutes are from time to time riven by internal disputes deriving from the way in which the direction of the larger institutes is the collective responsibility of senior people — the radioastronomy institute at Bonn seems to have been unlucky in this respect — the most obvious danger is that a group of people who age together will become less sparkling than at the outset of their enterprise. Proof is necessarily hard to come by, although Dr Dieter Raft, general secretary of the Max-Planck Gesellschaft, acknowledges that "we are not satisfied".

Most complaints against the institutes, however, are concerned not with the quality of the work but with the effects of the institutes on other research organizations. By offering security and freedom from teaching to talented younger scientists, does the system unfairly deprive the university system of talent? And should not the institutes be more closely integrated with their local universities? In reality some Max Planck institutes are closely integrated with universities, others are unnecessarily separate (see page 274).

0028-0836/82/210268-01$01.00

Money to spend

By any criterion, West Germany spends more on research and development than any other state in Western Europe. But the question keeps cropping up, "Are we getting value for money?"

Telling who pays how much for what, however, is complicated by the division of financial responsibility between the *Länder* and federal governments. The *Länder* governments are wholly responsible for the direct support of university research through recurrent budgets, contribute 50 per cent to the costs of DFG and some Max Planck institutes but only 10 per cent to the cost of large research establishments. Industry, of course, is separate.

In 1981, the total spending on research and development is estimated to have been just over DM41,500 million, something like DM750 per head of the population. The absolute amount is greater than that spent on research and development elsewhere in Europe. *Per capita* spending is close on twice the annual spending on research and development in Britain and, with the exclusion of defence research, more than is spent by government and industry in the United States.

Industrially financed research, mostly in industrial laboratories, accounted last year for more than half the total — DM22,490 million. Some officials suspect that this figure may be inflated. Government support for industrial research is estimated at more than DM5,000 million, mostly from the federal government.

There is similar doubt about the contribution of the *Länder* governments to research in universities, this was estimated last year to be DM5,380 million, but this is to some extent to be based on notional divisions of university expenses between teaching and research.

Among the government sources of finance, the federal ministry of research and technology (Bundesministerium für Forschung und Technologie, or BMFT) is dominant. This year (1982) the ministry had DM6,578 million to spend, apparently a real increase over the last year's DM6,073 million. But this year's total includes DM280 million for special programmes of research support in microelectronics, iron and steel and optical communications technology. Thus the funds available for continuing last year's programmes will increase less quickly than inflation.

BMFT is the source of support for most high technology in West Germany. Nuclear energy remains the biggest drain on the budget, with space activities and information technology of growing importance. More significantly, the technology ministry is the channel for the federal government's support of the Max Planck Society.

Centre of confusion

Argument about the future of the West German Cancer Research Centre at Heidelberg continues. The next landmark in the laboratory's turbulent contemporary history will be the special meeting of the governing body (the *Kuratorium*) arranged for 21 June, when the staff of the laboratory plans to reply to the criticisms of the report by the investigating commission under Sir Michael Stoker, published at the end of March (see *Nature* 8 April, p.481).

Professor Otto Westphal, previously director fo the Max-Planck Institute for Immunobiology at Freiburg, and director of the *Krebsforschungszentrum* since March this year, nevertheless hopes senior members of the staff will keep a pact he has made with them that there will be no public discussion of events of the past few years until after the *Kuratorium* meeting.

He does, however, say that the Stoker report was in some ways superficial, and too dismissive of the political nature of the problems that have plagued the centre. Referring to the resignation of his predecessor, Dr Hans Neurath, last summer, he says that it is now plain that there were faults on both sides — Neurath looked for too much change too soon, but was not adequately backed by government officials.

The cancer research centre has willynilly been caught up with politics because both the federal and the *Länder* governments have looked to it for advice on questions raised in the *Bundestag* and elsewhere on the incidence of cancer in West Germany. In the absence of formal channels for these consultations, differences of scientific opinion at the laboratory have been magnified into political differences — a circumstance further complicated by the election of a Christian Democrat government in Baden Würtemburg.

Westphal says that the effectiveness of the cancer research centre has hitherto also been impaired by a poor working relationship with the university clinic, part of which is in the next concrete and glass block. He says that there is already a more constructive relationship between the two establishments, partly because of the openness of some of those recently appointed to the university's medical faculty.

On the centre's role in medical research, Westphal says that West Germany is surely big enough to support a centre able to understand and acquire skills in new methods of treating various forms of cancer elsewhere in the world. Thus, he says, the complaint that the laboratory is not a centre of excellence across the board does not mean that it should be scrapped. In any case, its recent record is far from discreditable, particularly the work on new chemotherapeutic schedules and methods of analysis for the detection of small doses of carcinogens.

Perhaps the laboratory's most valid boast is the use of advanced image processing techniques for the accurate three-dimensional mapping of tumours (using some form of tomography), followed by the design of radiation sources (based on a specific implant of radioisotopes) which are optimized to destroy, say, a brain tumour but not surrounding tissue. The centre has the resources to deal with three such cases a week but depends on Heidelberg's neurosurgeons for implantation.

The somewhat down-to-earth character of those achievements is entirely consistent with the original concept of the institute by Professor K.H. Bauer, himself a Heidelberg surgeon and a personal friend of the first West German Chancellor, Konrad Adenauer. The hope in the 1950s was that applied research could help with the treatment of cancer. At first constituted as a foundation (*Stiftung*), the *Krebsforschungszentrum* became one of the research centres supported by the Ministry of Research and Technology as recently as 1976.

The centre's immediate difficulty is shared with the other major research centres dependent on the federal government — staff has to be cut across the board by 7.5 per cent in the next five years. If it is eventually decided that the pattern of the centre's work should be radically changed, the specificity with which the proposed cuts have been defined (so many scientific posts, so many cleaners) and the fact that half of the centre's original research groups are still in being will be a hindrance.

Westphal's own future is unclear. He relishes having brought a measure of calm to the laboratory at which, only a few months ago, one institute director (division head) told his boss at a committee meeting that he would fight him until one or the other went. (In the end, both went, the subordinate after an incriminating letter had been posted to Heidelberg in the wrong envelope.) It is a fair guess that Westphal will be looking for a measure of independence and an assurance of government support.

For the federal government, the problem is more serious. The *Krebsforschungszentrum* fits awkwardly in the clutch of major research laboratories financed principally by the Ministry of Research and Technology. (Its closest relative is the biotechnology laboratory at Braunschweig, west of Hannover.) In reality, however, the existing federal grant-making agencies are conspicuously unsuccessful at supporting clinical research — some say because physicians are so busy making money. With an election two years or less away, will the federal government have the stomach for the long process of consultation, with the *Länder* and everybody else in sight, that might eventually lead to an effective medical research organization? □

0028-0836/82/210269-03$01.00

Particle physics goes ahead

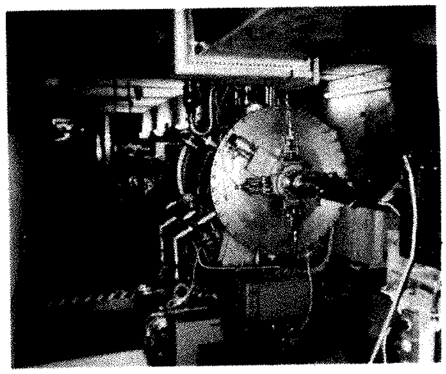
BEFORE the Second World War, German electron physics and technology was already strong. And despite the flight of physicists from the Nazis, it remained strong afterwards, partly through the influence of the "father of quantum mechanics" Werner Heisenberg and his colleague Heinz Maier-Leibnitz. The tradition has been maintained in one place in particular: the Deutsches Elektronen-Synchrotron in Hamburg.

DESY, as it is called, has become the only national high-energy physics laboratory in Europe which is unquestionably of world class. So much so, that a couple of years ago there was a distinct possibility that DESY would upstage CERN, the big joint European high energy laboratory near Geneva, and go ahead independently to build the next big machine for Europe (the 27-km circumference large electron–positron collider, LEP).

How has DESY managed to become so strong? The laboratory owes much to the canniness of DESY directors in raising money within Germany; and also to good, cost-cutting design of accelerators, rather in the legendary style of Fermilab's Bob Wilson in America. It is not exactly the spirit of Göttingen again: more like German technology and good management. Also it has been luck. DESY is an electron laboratory, and electron physics has come to the front of the stage, as the electron is now seen to be elementary, in contrast to the other commonly accelerated particle, the proton, which contains three quarks. DESY got into the excitement of charm, and the *Land* of Hamburg (which contributes 10 per cent of DESY's budget) and the federal government in Bonn now look on the place as a little German jewel. The previous director, Professor Herwig Schopper, whose political lobbying in Bonn and Hamburg was so effective, is now director-general of CERN, but his successor, Professor Volker Soergel, although a more private and very different man, has taken over all of Schopper's tactics — in particular, internationalism.

Soergel has attracted the largest American high energy physics experiment ever to be undertaken in Europe: \$4 million of equipment and 30 American physicists will stay at DESY for upwards of three years to make a thorough investigation of the interactions of the bottom quark (the one that came after charm). According to Elliot Bloom, the Stanford University spokesman of the group, there is "tremendous support" at DESY. Bloom hopes to match and perhaps outgun the only equivalent "bottom" investigation in the world, at the CESR facility at Cornell University, New York.

Bloom will use as the collider DORIS, the electron–positron collider that first put DESY on the map during the charm era of 1974–76; but now it has been completely rebuilt, and should do for the higher mass of "bottom quark" what the old DORIS (and the Stanford machine SPEAR) did for charm.



Superconductivity success at DESY — an accelerator cavity which works on a storage ring.

The other big machine at DESY is PETRA, which shares with PEP at Stanford the accolade of being the highest energy electron–positron collider in the world. However, both PEP and PETRA

More neutrons may cost too much

Is neutron scattering "Big Science"? It is at a price of DM850 million (about £220 million), which is the current tag on a spallation neutron source (SNS) being proposed in West Germany. The idea is to produce a high flux of neutrons, particles which can be used rather like X-rays to analyse the structure and dynamics of many kinds of material, by bombarding a target of uranium with a pulsed beam of protons. Broadly speaking, fluxes much greater than those currently available with "high flux" nuclear reactors can be created in that way, so spallation neutron sources are now seen as the second generation sources in the field, just as synchrotrons have replaced X-ray tubes in X-ray work.

However, spallation sources are not cheap. Only HERA, the electron–proton collider planned by the national high energy physics laboratory DESY (see above), comes close to the SNS in price among major capital projects currently before the ministry of research, and even that is cheaper. But the SNS would provide for a much larger community of scientists than HERA, and it could be built in two half-price stages. Which to buy, HERA or the SNS? Or is there money for both? The decision is due sometime next year.

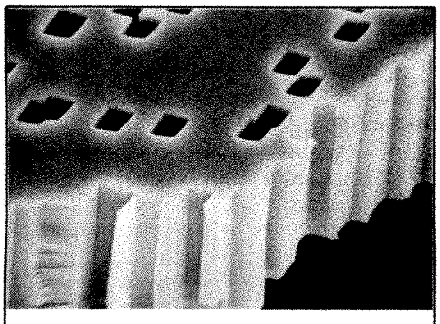
For the moment, the two interested communities are playing a kind of poker game. The minister for research and technology, Dr Andreas von Bülow, has asked them to assume that the government would fund both projects, but to stretch the construction schedules of their machines so that the net annual cost would not be too high. However, von Bülow has omitted to say what "too high" would be; so the communities are now competing both against each other and against an uncertain ceiling to pare their costs to the minimum. Nevertheless there is some hope that a recent government commitment to make a major investment in the economy in the middle of next year (at the expense of increased borrowing) may see money being spent on both projects, if only to help the construction industry. This was just how DESY first got the money to build PETRA (see above) some years ago.

The only firm decision so far is that the SNS, if funded, will be built at the Jülich nuclear centre near Cologne, where Professor Hans-Heinrich Stiller is the project leader. (Karlsruhe, an early competitor for the site, has been eliminated.)

The SNS would be bigger (and come on line considerably later) than the only other similar machine in the world, the British SNS under construction at the Rutherford laboratory — which will first become operational in 1984. The German SNS would not be likely to be ready before 1991, even if it got early approval; but then it would produce more than three times as many neutrons (peak flux) as the Rutherford SNS in the first stage, or 25 times as many in the second stage. The corresponding figures for mean intensity are 25 times and 200 times the British SNS — an indication of how sophisticated and complex the German technology would have to be.

On both the British and the German sides the open question is how much collaboration can be established on the two projects. Money for the British SNS is agreed, but a lot more cash is needed for the complex instrumentation (spectrometers and so on) which must accompany a neutron source if it is to be a success. (A major reason for the pre-eminence of the high-flux reactor, the Institut Laue-Langevin at Grenoble, in this field can be traced back to an early and effective emphasis on instrumentation for the institute.) So Britain would be happy if Germany could contribute a few instruments to the SNS. Then, perhaps, if the British economy had improved, there might be some British support for the German machine in the 1990s. Very preliminary discussions are already under way on such matters. But a likely fate for the large-scale German SNS might be that it will not be built at all, unless much wider international collaboration can be set up to support it.

Robert Walgate



A molecular sieve (see article on GSI, right).

have suffered a trick of physics: the energy region has proved so far to be relatively barren of spectacular results. Nevertheless, PETRA has found evidence for gluons (the photons of the interquark force), and has more recently demonstrated "interference" between the electromagnetic and weak interactions — which indicates the existence of the neutral intermediate vector boson, the present holy grail of particle physics.

Another DESY success (in collaboration with Karlsruhe) was demonstrated only this month: "our little present for LEP", as it is described at DESY. It is a superconducting accelerator cavity, which can reduce the electric power required to drive an electron accelerator by a very large factor (up to tenfold) or, equivalently, raise the energy that the machine can reach for the same power input. Since CERN's likely power bill for LEP is a major headache, the demonstration of working, cheap and reproducible superconducting cavities would be a great step forward. Two weeks ago a single Karlsruhe-DESY cavity successfully drove PETRA to 5 GeV. The cavity has some unique features, and could be produced industrially, but so far it is too expensive: to equip PETRA fully with the devices would cost DM50 million, DESY staff estimate, for a present electricity bill saving of DM2–3 million a year. Nevertheless, savings at LEP would be much greater, and it is clear the design is approaching being practical.

DESY is again looking abroad to raise the finance for HERA, DESY's big (DM600-million) project for the 1990s. (HERA would collide 30 GeV electrons with 820 GeV protons, to probe the lepton–quark interactions. It would use superconducting bending magnets. If for some reason LEP failed, it could be converted to create the intermediate vector boson.) A DESY spokesman said recently that negotiations are already well-advanced if not with governments, with laboratory directors elsewhere, with hope of raising around DM200 million of the cost of HERA abroad. Hamburg has already committed itself to its 10 per cent (DM60 million) which would then leave Bonn with only DM340 million to find — cheaper than the first stage of the spallation neutron source with which HERA might be in competition for Bonn resources (see opposite).

Robert Walgate

An element of risk

The Gesellschaft für Schwerionenforschung (GSI) at Darmstadt (south of Frankfurt) is riding high. It has one brilliant feather in its cap — the artificial production of element 107, the heaviest so far. And its director, Professor Gisbert zu Putlitz, has ambitious plans for bigger and better experimental equipment.

GSI, the only dedicated heavy-ion laboratory in Europe, lives by its equipment for accelerating heavy ions — even uranium ions — in several stages. At the end of last month, the laboratory inaugurated an extension to its UNILAC accelerator that will double the maximum energy of the ion beams produced to 20 MeV per nucleon. But zu Putlitz is already looking beyond that target.

Element 107, represented by the isotope with 155 neutrons, was made at GSI last year by the collision of accelerated chromium-54 with a target of bismuth-209. As with all experiments of this kind, the chief difficulty is to demonstrate that the intended element has indeed been produced. The team at GSI used a novel arrangement of electric and magnetic fields to demonstrate their few nuclei of element 107 and thus to break the near-monopoly on the creation of heavy elements which the Lawrence Radiation Laboratory in the United States has enjoyed in the past few years.

More intriguing, however, is the demonstration of a novel form of radioactive decay — the emission of protons rather than electrons from nuclei that are deficient in neutrons. By the collision of a beam of nickel-58 ions with a target of ruthenium-96 for example, it has been possible to synthesize the isotope lutetium-151 which has 24 neutrons fewer than the stable isotope of the element. (The isotope thulium-147 is also subject to proton decay.) Plainly this novel form of nuclear instability has been a great fillip to nuclear physics — and not merely in West Germany.

The laboratory says, however, that it is not merely a nuclear physics establishment but a practical place. By way of evidence, it cites the use of molecular beams for producing molecular sieves. What better than a beam of heavy ions to punch holes of molecular dimensions in a synthetic membrane? So filters are being made for instruments that measure particulates in air and models constructed for studying the permeability of biological membranes.

So what of the future? Professor zu Putlitz clearly regards the present energies at which heavy ions can be produced as only a modest beginning. Ultimately, there will be a need for ions with energies of 500 MeV per nucleon or even 10 GeV per nucleon. Such energies cannot be achieved with a linear accelerator of reasonable cost and size, and GSI has therefore proposed to build two interlinked synchrotrons, presently called SIS 12 and SIS 100. The smaller one, SIS 12, would use UNILAC as an injector and would accelerate ions to about 1 GeV per nucleon. The next synchrotron (SIS 100) would take over from there and further accelerate the beams to the maximum energy of about 10 GeV per nucleon.

Enough land is available for both machines. The cost, just under DM90 million for SIS 12 and just over DM 150 million for SIS 100, is the stumbling block. But SIS is high on the government's list of major projects; zu Putlitz is optimistic. Indeed, if funds are short, he would prefer to start by building SIS 100. Time alone will tell whether he will be able to follow this more imaginative but no doubt risky approach.

Konrad Guettler

Research without responsibility

Professor Klaus Pinkau, director of the Max Planck Institute for Plasma Physics at Garching-near-Munich, is a square-jawed square-built man with a shock of grey hair and a ruddy complexion who exudes enthusiasm. He thinks he knows something of what is wrong with the state of science in West Germany.

First, he says, people are needlessly depressed by what is in effect an illusion, what he calls the "echo-effect". Everybody must accept that the United States is now the chief source of scientific discovery, and American scientific journals the chief medium for the announcement of new developments. Inevitably, however, this implies that papers by West German scientists are referred to less frequently, and less fulsomely, than their merit would demand. Worse still, there is a tendency for science journalists and even scientists themselves to take note of some development in West Germany only if it is first applauded, or approved of, in the United States.

In reality, Europe (and West Germany) has more to boast about than anybody acknowledges. Has not the COS B gamma-ray satellite (switched off last month after seven years of operation) been a unique and successful demonstration not merely of the technique but of the capacity to formulate an adventurous project?

So what else is good in West Germany? Pinkau echoes the common response by referring, first of all, to "Eigen's group in Göttingen". But his main impression is that West Germany has done best in those

0028-0836/82/210271-02$01.00

fields in which there were enough people still active after the Second World War to provide a nucleus of active research. Perhaps nuclear physics, especially the work with heavy ions at Darmstadt (see page 271), is representative of the best.

Professor Klaus Pinkau: "We do second-rate research rather well".

Like others, Pinkau thinks that bureaucracy has had a baleful influence. Why should it have been left to a committee of the *Bundestag* to decide, last December, which six posts at his laboratory should be abolished? But the irritation caused by such happenings is less important than the way in which the bureaucracy's existence undermines the sense of responsibility that professional scientists should enjoy. If everybody knows that, in the last resort, distant committees of officials will make the most detailed decisions, how can working scientists shoulder the responsibility implicit in their promises?

Pinkau shares the common concern about the reform of the universities, but in a novel way. The problem is that the universities have also lost their sense of responsibility. He agrees with the opinion that "we do second-rate research rather well". He is doing what he can, personally and by encouraging his colleagues to follow his example, by teaching regularly at the technical university at Munich.

The lack of responsibility — the freedom to act independently — is here again Pinkau's main worry. "The government", he says, "must realize that by restricting the independence of scientists, they are killing science". He acknowledges that the Max Planck Society has an enviable reputation for not being bureaucratic, although his own institute is in the significantly different position of being financed directly from Bonn.

But even with this proviso, Pinkau answers a firm "yes" to the question "Should the Max Planck institutes be merged with their neighbouring universities?". That would be good for research, but in any case there is too much tension between university and Max Planck scientists for comfort. □

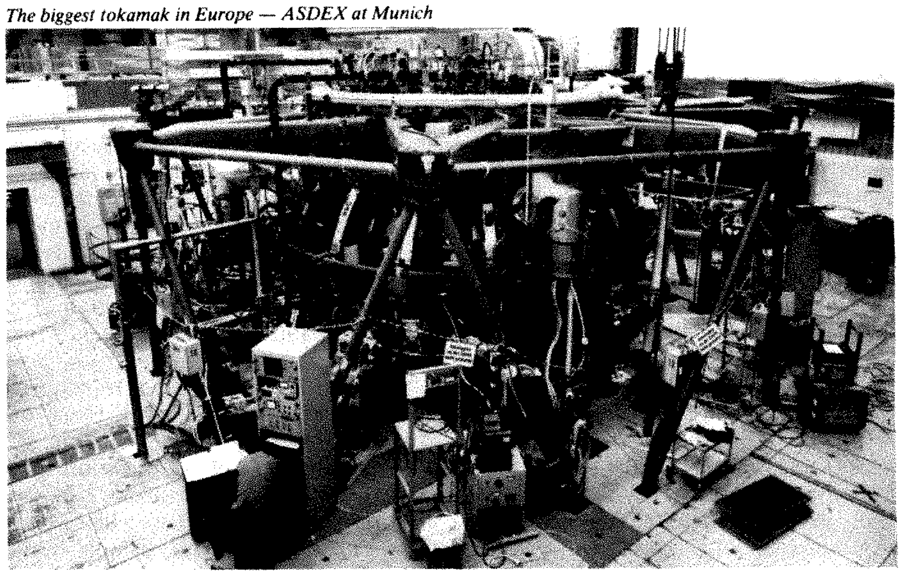
New ways with fusion

In 1978 the European decision to build the Joint European Torus (JET) at Culham in the United Kingdom rather than at Garching in West Germany seemed a blow to the German plasma physics community. For JET is the principal experimental facility in the Euratom plasma physics programme, and for when national programmes have to compete for funds with collaborative international experiments whose funds are necessarily committed in advance, there is always a danger that they will be squeezed.

The main plasma physics research laboratory in West Germany, the Max-Planck-Institut für Plasmaphysik (IPP) at Garching near Munich nonetheless continues to run a thriving national programme. But the generally less favourable economic climate of the past few years has taken its toll; an initial multitude of experiments has gradually been reduced to two main lines of research. One of these is centred around ASDEX, the biggest tokamak machine in Europe, which was commissioned in 1980 and has already yielded important scientific results.

What excites the people at Garching about ASDEX are the diverter coils which produce a magnetic field allowing for control of the level of plasma impurities. The stability and lifetime of any contained plasma is severely limited by the impurities introduced through interactions between the plasma and the wall of the container. The diverter chambers of ASDEX significantly reduce the level of metallic impurities and lighter species such as

The biggest tokamak in Europe — ASDEX at Munich



Jülich's jewel

"Come and see our jewel", said the man behind the wheel, pulling up outside a hanger. The jewel in question is the fusion experiment called TEXTOR, a machine recognizable as what it is from its toroidal shape but with the advantage that it can be separated into two halves so that the interior can be rebuilt.

TEXTOR, partly financed by the European Community as part of the joint programme of thermonuclear experiments, has so far cost DM110 million, with close on DM50 million still to go. With its discreet constructor's label, Krupp, the machine is as nicely engineered as it could be.

The torus is intended for the study of problems that may crop up in thermonuclear reactors thirty years from now. It is a small tokamak designed to examine the interaction between a hot plasma and the containing walls, the behaviour of wall materials under neutron irradiation and the processing of tritium, the necessarily radioactive component of likely thermonuclear fuel. Everybody is delighted with the way it works.

For a nuclear research centre such as the Kernforschungsanlage Jülich, set up in 1956 by the government of Nordrhein-Westfalen but since 1970 supported to the tune of 90 per cent by the federal government, a long-term project such as this must be a comfort.

The establishment is best known for its development of a high-temperature reactor, an experimental version of which went into operation in 1967, producing 46 MW of heat. The prototype of that design being built at Hamm-Uentrop nearby will function only in 1985 (seven years late), producing 300 MW of electricity.

The technical problems are those of making reactor components that will operate reliably at 950°C for years on end, but the costs are also high. So Jülich is pushing hard a scheme for long-distance energy transmission in which fast reactors would convert methane (and water) into hydrogen and carbon dioxide for reconversion at some distant point.

0028-0836/82/210272-02$01.00

oxygen. The mechanical devices used elsewhere in tokamak machines are said to be comparatively ineffective.

The second objective of ASDEX is to study the properties of plasmas heated by beams of neutral atoms. The original neutral beam injectors are being upgraded to a total power of 2.5 MW which should lead to an increase in plasma temperature from 6×10^6K to 2×10^7K. There are plans to use radio-frequency heating to increase plasma temperatures still further, and to use a deuterium pellet injection system to replenish the plasma particles removed by the diverters.

The ASDEX device will remain the major focus of research for several years, but plans for an upgraded version are already being drawn up with support (in the design phase) from the European fusion programme. As well as being a bigger machine, the distinguishing feature of the version being planned is that the diverter coils will be external to the main toroidal field coils, an arrangement considered to be essential for a fusion reactor.

Most of the research at ASDEX is of direct relevance to JET (also a tokamak machine) and the laboratory says that there are close links between the two establishments. But IPP has not put all its eggs into one basket; it continues its research into the physics of stellarator machines. In the present machine (called Wendelstein VII-A) the peculiar stellarator magnetic field configuration is achieved by a set of toroidal field coils surrounding the inner helical coils. But the laboratory is planning an Advanced Stellarator in which these two sets of coils would be replaced by one, each of them with a somewhat different twisted geometry. The result is expected to be improved plasma stability at a higher plasma pressure. The design phase has begun and the new machine could be working as soon as 1985.

Although the Garching institute is technically part of the Max-Planck-Gesellschaft, its size and large annual (running) budget of DM110 million made it a natural candidate for membership of the association of the large science centres, AGF. The institute's finances are outside the usual Max Planck pattern — a quarter of the budget is derived from Euratom, with the remainder split 90:10 between the federal government and the *Land* of Bavaria.

Not surprisingly, other Max Planck institutes are uneasy at the presence of such a cuckoo in their nest. Fusion research is, however, one of the few European programmes that really work, and it is difficult to imagine how such a successful research centre could be effective on a much smaller scale. Nevertheless, as the size of fusion reactor experiments becomes even larger in the future, it is likely that one of the more applied research centres, such as the huge *Kernforschungszentrum* at Karlsruhe may take a bigger share of the development — its contribution to research at Garching takes the form of technological development of such items as the pellet injection system.

Within the national German fusion programme, the Karlsruhe centre tends to be most active in magnetic field technology and materials study, but the centre is also involved in system design studies. In particular, it maintains a modest presence in two lines of fusion research: inertial confinement fusion using light ion beams and fusion by magnetic confinement in a tandem mirror machine. The light ion fusion studies involve a pulsed high-power ion generator, whereas the tandem mirror configuration is only at the system design stage. Earlier mirror experiments at Lawrence Livermore National Laboratory in California yielded promising results, yet the research at Karlsruhe seems to be the only investigation of its kind in Europe.

Fusion research in West Germany seems thus to have retained an enviable diversity, the commitment to JET notwithstanding. The machines now operating, at IPP and elsewhere, are producing a substantial harvest of data. And with funds secure for the initial upgrading of the major facilities, the institute has a viable programme for the 1980s. Its involvement in the study groups for the Next European Torus (NET) and the international INTOR project should leave it well placed in any competition for a future international facility.

Yet the time scale for break-even fusion power seems to be receding further and further; Klaus Pinkau, director of IPP, is quite frank in saying that two or three steps will be necessary after operating experience with JET, perhaps by the end of this decade. Each step seems to require ten years or so, so that it may not be known until the year 2020 whether fusion is technically and economically feasible. "I'll be dead by then", he says.

Konrad Guettler

Alternative routes to fusion

THERE are other means of inertial confinement fusion which have received considerable attention and funding elsewhere: laser fusion and heavy-ion beam fusion. The main difficulty with the former is the high power output required by the lasers and reliable operation with a high pulse repetition rate. Heavy-ion beam fusion research suffers from the handicap of requiring a very large initial investment in the ion beam accelerator. Some of the relevant ion beam studies are being carried out at the GSI accelerator (see below) near Darmstadt but with low priority. A high-power laser project group originally formed part of IPP at Garching but some time ago it was expanded and elevated to the status of the separate Max-Planck-Institut für Quantenoptik. Although still sharing a site with IPP at Garching, the research emphasis of the new institute is divided between laser fusion, laser spectroscopy and laser chemistry. Its fusion research is carried out with a high-power iodide laser and is aimed at understanding the fundamental aspects of energy transport within the compressed fusion pellet. The emphasis on some restricted topics of research appears to be deliberate, although some foreign concern seems to have been expressed about a much more expanded national laser programme.

Konrad Guettler

An astrophysics star is rising

WEST GERMAN astronomy and astrophysics are neatly parcelled up on wavelength lines: the Max-Planck-Institut für Radioastronomie in Bonn (see next page) looks after the radio-wavelength part of the spectrum, the Max-Planck-Institut für Astronomie in Heidelberg looks after the optical region and the Max-Planck-Institut für Physik und Astrophysik (with its Institut für Extraterrestrische Physik) in Munich deals mainly with the short wavelength domains of UV, gamma-ray and X-ray astronomy. The Munich institute is the major West German institute involved in satellite-based experiments and has an ambitious programme for the 1980s.

The gamma-ray part of the package is perhaps the most distinctive. Balloon-borne telescopes are used for low energy studies, while the high-energy domain has been studied as part of a European collaboration using the COS-B satellite, shut down only in April this year after more than six years of operation. The users are delighted with the way this instrument has performed but also with its conception. "It shows", Professor Joachim Trümper, head of the institute, says, "that European scientists can still be the first with a good idea".

Even so, the prospects for further research in the field now depend on plans for the Gamma-Ray Observatory (GRO) satellite of the US National Aeronautics and Space Administration (NASA) for which the insitute is helping to design one of the telescopes. Its other gamma-ray investigations centre mainly around the Solar Maximum Mission and the ISEE-3 satellites, which are specially equipped to study solar flares and gamma-ray bursts.

The magnetosphere and the interplanetary medium constitute the second major focus of the institute's programme. The Earth's magnetosphere is being

0028-0836/82/210273-02$01.00

investigated with experiments aboard the European GEOS-2 satellite, and the interactions between the magnetosphere and the solar wind with the ISEE-1 and ISEE-2 satellites respectively. While the analysis of the data from these satellites will keep people busy for many years, what excites them most are the preparations for the next mission — a major joint project between NASA and European institutes called the Active Magnetospheric Particle Tracer Experiment (AMPTE). It consists of three major components. The Munich institute is building an Ion Release Module, NASA is responsible for the Charge Composition Explorer satellite and the British Science and Engineering Research Council is contributing a subsatellite equipped for *in situ* plasma diagnostics. The release module will inject artificial plasma clouds into the solar wind and the tail of the magnetosphere, and the two other satellites will measure the interactions between the cloud and the ambient plasma.

For the future, the X-ray astronomy group seems to have even more dazzling prospects. Apart from its own balloon experiments, the institute is a major participant in the EXOSAT satellite soon to be launched by the European Space Agency. But its most ambitious project is ROSAT, a large X-ray observatory designed for launch by the space shuttle in 1986 or 1987. This instrument has been designed to have much better angular and spectral resolution than even the NASA Einstein Observatory which ceased functioning last year.

Primarily a national project, this satellite should give West German scientists a commanding place in X-ray astronomy at the end of the decade. The initial observational programme is a sky survey for the first half-year after launch which should reveal many more X-ray sources, given that the satellite's sensitivity is three times greater than that of the Einstein Observatory. This survey is to be followed by a one-year study of interesting sources. If all goes well, the satellite could even be retrieved after a couple of years by the space shuttle and refurbished.

Although a national project, ROSAT will provide observation time for other astronomers. Negotiations are under way with NASA for a US instrumental contribution in exchange for time, while there is to be an extreme-UV telescope built by a consortium of British universities.

So in the late 1980s the focus of attention in the field will shift from America to Europe. Repeated cuts in the US research budgets have put the competing Advanced X-ray Astronomy Facility further back, even though it has received the highest priority in the recent Field report (see *Nature* 8 April, p.482). And with no successors in sight for the UK Ariel satellites, ROSAT will be the only source of X-ray data for years to come.

Konrad Guettler

Contrasts in Max-Planck styles

MAX PLANCK INSTITUTES seem to come in all kinds of shapes and sizes. The largest and most costly of them is the institute for plasma physics at Garching, but the radioastronomy institute at Bonn is similar in that it owes its existence to equipment so expensive and distinctive that it cannot easily be the property of a single university department. The institute's boast is a 100-metre high-frequency radiotelescope.

At the other end of the Max Planck spectrum is the newly created (1980) Institute of Mathematics, which is closely a part of the University of Bonn and which has the further distinctions that its personnel consist largely of visiting mathematics professors (with graduate students from the university) and that it is the first example so far of a Max Planck institute created from one of Deutsche Forschungsgemeinschaft's special research programmes (*Sondersforschungsbereichen*, see page 266). Although the institute will not become a charge on the Max-Planck-Gesellschaft until 1985, it already has a shiny new name-plate on its handsomely constructed building.

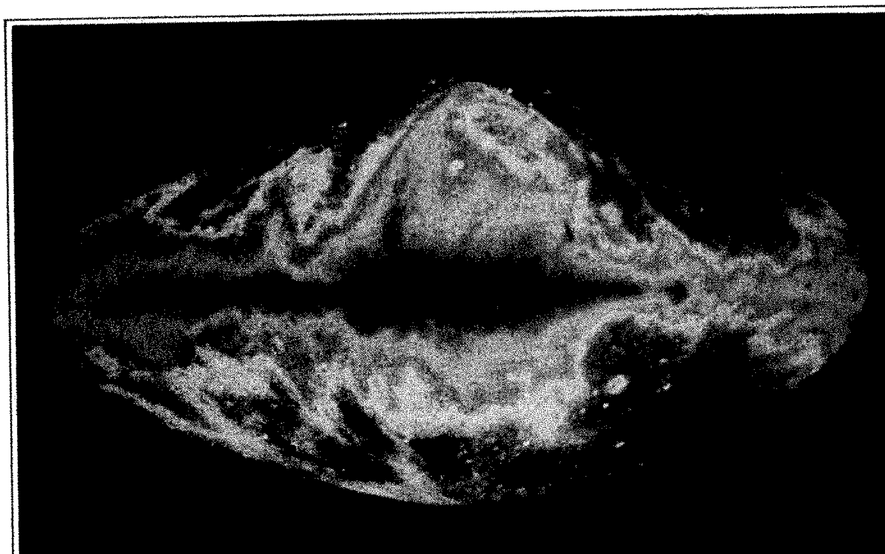
Technically, the radioastronomy institute is inevitably the more spectacular. The 100-metre dish sits in a well-wooded valley at Effelsberg, 30 kilometres east of Bonn. It has been designed to operate at frequencies up to 50 GHz (a wavelength of 6 mm) and beam width at 1.2 cm is reckoned to be 35 seconds of arc.

The total cost of the dish (completed in 1972) and the ancillary equipment was DM 40 million, a substantial part of Max-Planck-Gesellschaft's capital budget more than a decade ago. The staff (including a score or so of visitors) amounts to more than 180 people, of whom a third are scientists. The institute is divided between an office in Bonn and the field station at Effelsberg, and includes a substantial technical effort at both places on the design and construction of detectors for use with the telescope.

Operationally, the telescope has been an immense success, at least in part because of the cleverness with which the structure was designed. The objective is somehow to preserve the parabolic shape of the 100-metre dish whatever the elevation. This has been accomplished in the Krupp-MAN built instrument at Effelsberg by attaching a huge counterweight by a network of struts to the back of the dish so that the dish can distort only into another parabola. The result is a dish that is substantially lighter than rigid construction would have required.

The staff at the mathematics institute is by comparison inconspicuous. The director, Professor Friedrich Hirzebruch, says that he and his two senior colleagues from the mathematics department at Bonn remain on the university's payroll, but that he is relieved of administrative duties (but not teaching) on account of the time spent at the institute. He thinks he may apply to Max-Planck-Gesellschaft for two permanent positions for mathematicians, but the bulk of the budget, roughly DM 2.5 million a year, will still be used to support mathematicians visiting from elsewhere in West Germany and abroad.

In Hirzebruch's view, an institute of mathematics is necessarily a place for peripatetic people. Since his group was formed in 1969, he says that it has acquired an international reputation comparable with, say, those of the groups at the Princeton Institute of Advanced Study, the



A map of radio-sky intensity at 408 MHz compiled at the Max Planck radioastronomy institute at Bonn from data gathered over a period of 15 years at Bonn, Jodrell Bank and Parkes (Australia). Data reduction was by means of a Cyber 172 computer at Bonn. Prominent patches along the mid-line of the sky are the local arm complexes Cygnus X (left) and Vela X (right).

0028-0836/82/210274-02$01.00

Institut des Hautes Études near Paris and the University of Warwick as being places where mathematicians can work on interesting problems with interesting colleagues without having permanently to tear up their roots at the place whence they came. But Hirzebruch also sees his peripatetic working group as a direct benefit to the mathematicians at Bonn.

The links between the institute of radioastronomy and the local university seem much less immediately productive. Professor Richard Wielebinski, the Polish-born Australian immigrant who is now one of the laboratory's two directors (the executive director is Professor Peter G. Metzger) says that the radioastronomy institute is, above all, a national research centre. A substantial part (more than 20 per cent) of the 100-metre telescope's time in the past two or three years has been used for coordinated interferometric observations in which radiotelescopes as far apart as California (Goldstone) in the West and the Soviet Union (Crimea) in the East have been used to produce the intensity of radio emission from distant objects such as the radio galaxy 3C298 with a resolution of 0.5 seconds of arc.

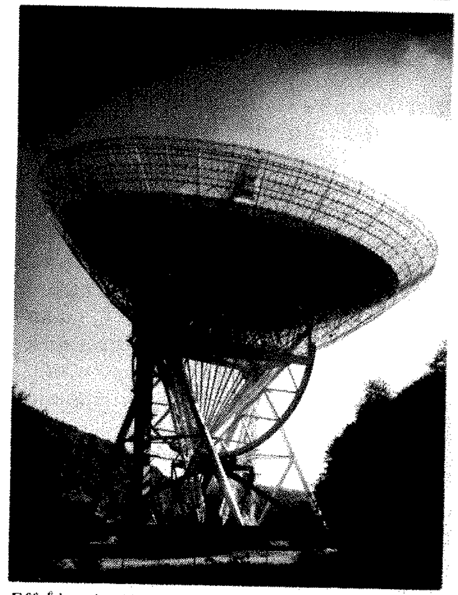
Long baseline operations such as these require long-term planning and the blocking-out in a somewhat arbitrary fashion of other opportunities for observation. Even so, the institute is planning to extend its international operations. The institute is the vehicle for the Max-Planck-Gesellschaft funds being channelled into the construction of a 30-metre radiotelescope being built in Spain to operate at frequencies of 80 GHz and above. This will be for a time, when tested in operation next year, the largest instrument of its kind in this region of the spectrum. The instrument will be used for observations of extragalactic carbon monoxide and continuum radio emission (the latter being one of Wielebinski's particular interests). It is also intended that the Spanish instrument should be used in conjunction with an array of three 15-metre dishes being built near Grenoble in France under a fifty-fifty agreement between Max-Planck-Gesellschaft and Centre National de la Recherche Scientifique, but that agreement has been made at headquarters level.

Off its own bat (but with approval from Munich), the institute has made a fifty-fifty deal with the University of Arizona that will lead to the design and construction of a 10-metre sub-millimetre telescope, to be built in Arizona. That agreement is to be signed in June this year.

What these two institutes therefore have most conspicuously in common (from 1985, in the case of the mathematics institute) is that the money comes from Munich. There is also a sense in which a direct comparison between them is unfair. One (the mathematics institute) is a device for channelling support for old-fashioned scholarship through a talented scholar to a university which, even though founded in 1818 in the aftermath of the Napoleonic wars, is distinguished in this and other fields. The other, the radioastronomy institute, is either a national research centre, intended to put West German radioastronomy on the map, or a national research facility, to be used by whichever West German astronomer has a sensible idea as to how the equipment should be used.

The trouble seems to be that this Max Planck institute pretends to be neither of those things but rather, in the Max Planck rhetoric, to be a free-living community of independent scholars which, nevertheless, by means of its senior people's honorary professorships (at the local university), its downtown location so close to the Department of Astronomy at the University of Bonn that the two institutions can share a library and its willingness to take in graduate students from everywhere, is intrinsically part of the university system. The irony is that the university's department of astronomy seems likely to succeed in its application to Deutsche Forschungsgemeinschaft for the funds with which to build another radiotelescope at a different site.

What follows is a strictly personal impression. First, the two institutes have distinctly different roles. The mathematics institute is not merely a source of scholarship but a way of enlivening mathematics at the University of Bonn, while the institute of radioastronomy, in spite of its links with university astronomers in West Germany and elsewhere, must be judged primarily by its own staff's output of research results. To some extent, of course, these differences are explained by the differences between the subjects — mathematicians are mobile, radioastronomers to some extent tied to their instruments. But there are other ways of organizing the use of large research facilities, with the American pattern of open access through user committees at one extreme and the British tendency to base radiotelescopes in university departments at the other. More concessions at Effelsberg in one direction or the other could not fail to embed the use being made of this splendid instrument more firmly in the scientific community's consciousness.



Effelsberg's 100-m dish: productive but an ivory tower?

Making a business of research

By what trick can a group of only four tenured academics run a department with 400 students, 160 of them doctoral students, spend the best part of DM20 million a year and yet keep the department's well established reputation for being the best academic institution of its kind anywhere in the world? The department is the machine tool institute at Aachen, one of the fourteen specialized institutes in the mechanical engineering complex at the Rheinisch-Westfälische Technische Hochschule (RWTH). The trick is to run the department as if it were a business.

The four tenured academics are university professors. Their formal duties are to lecture to the undergraduates, but each also has under his wing a small clutch of research groups, perhaps three or four. The research groups are run by senior engineers, hand-picked from among those gaining a doctorate in the department and employed as the need arises on short-term contracts.

Of the total cost of the department, no less than eighty per cent comes from outside RWTH, principally from the federal government and its grant-making agencies. The brunt of the research work falls on the graduate students, paid out of project grants, but the department also makes a special point of engaging undergraduates part-time in its research work (and pays them for their trouble). Ten per cent of the successful diploma candidates are kept on for doctoral work, on which they spend between five and eight years (the average is about six). Once their dissertations are completed, they must find a job elsewhere.

That step appears to present very little difficulty. Apart from the institute's high reputation, the Laboratorium für Werkzeugmaschinen und Betriebslehre prides itself on having the closest possible relations with companies in the metal-forming industries of West Germany, many of which play a crucial part in helping to define the department's research programme. But the department will add a project to its portfolio (and apply to some

0028-0836/82/210275-02$01.00

Hard management and software

THE machine tool institute's reputation stems directly from the single-mindedness of its founder, Professor Adolf Wallichs, who in 1906 seems to have taken fright at developments in production engineering in the United States and to have embarked on a research programme in a small cellar. One of his achievements was to persuade German industries to contribute to the cost of a building and the equipment to go with it at the height of the post-war inflation in 1924.

Wallichs was succeeded in 1938 by Professor Herwat Opitz, whose 37-year spell in office saw radical developments in, for example, the design of gear-cutting tools and, after the Second World War, the threefold growth of the institute. Since 1973, the management of the institute has been shared between the occupants of the three full university chairs (*Lehrstuhlen*) among whom the chairmanship of the department rotates. (Metrology is for the time being the poor relation.)

The research programme is now dominated by the application of computers to production processes. Thus the design section is up to its eyes in computer-aided design (CAD). Here and elsewhere, part of the objective is to produce a library of software that can be used by the small to medium sized manufacturers that dominate this sector of West German industry in the optimization of, say, a gear train — and the automatic draughting of the drawings needed. Similarly, using banks of data built up over the years on the characteristics of various cutting and metal-forming processes, computer programs are being designed so as to identify the best way of producing metal parts of various forms.

The search for accuracy and reproducibility is a recurrent theme in the research programme, whence the current interest in real-time measuring techniques. Similarly, computer techniques are being used in the analysis of how machine tool structures are deformed by stresses arising during use.

While all this software is accessible to would-be industrial users, specialized seminars (at Aachen but also at Zurich) are regarded as essential means of propagating the new ideas that arise, as are the biennial international conferences on machine tools, the last of which last year attracted 2,000 participants.

such organization as the Deutsche Forschungsgemeinschaft for funds) only if the outcome of its work can be published in the open literature. Consultancy on specific problems for industrial companies is only a modest part of the department's work.

This unique way of doing academic business has, in effect, made the institute both the research laboratory and the source of skilled engineers for much of an important sector of West German industry. Yet this is one of the smaller of the fourteen institutes catering for the 5,300 mechanical engineering undergraduates now registered at Aachen.

But is this not a narrow way of training engineers? Yes and no is the customary answer. For the first two years of their academic careers, students follow a more or less common course, heavy on mathematics and physics or chemistry as the intended speciality may suggest. On paper, the undergraduate course should take four years, or eight semesters, divided into the two equal halves of *Grundstudium* and *Hauptstudium*. In reality, par for the course is more like six years, although a small percentage of students win their diplomas after 4½ years. Along the way, some 30 per cent of Aachen's engineering students fail to jump one hurdle or another, although it is reckoned that only seven per cent of them fail to complete some kind of first-degree course, perhaps after moving to some other university. Hothouses like the machine tool institute are responsible only for the second half of an engineer's education.

The rigours of Aachen's engineering courses seem not to have deterred school-leavers from signing on. Indeed, business is booming. This year's intake in engineering is up on 1980–81 by between 10 and 20 per cent, according to the field, while the volume of enquiries suggests that there will be a further jump next October. The proportion of women remains negligible.

The apparently steady increase of the length of time needed to win the right to the title *Dipl.-Ing.* is nevertheless a source of some concern. There is an unavoidable conflict between the need to give graduates some marketable breadth in their fields and the objective, still held to, that graduates should know about everything there is to know.

The decline in the knowledge of entering students of mathematics and basic science has also made it necessary for RWTH to provide extra courses in mathematics in the September before the first year begins. Apparently what has happened is that students seeking high marks in the school-leaving *Abitur* are opting out of the extra three hours of mathematics offered each week in the last two *Gymnasium* years.

In this respect, RWTH makes more concessions to its students' needs than many other West German universities, where students are usually left largely to their own devices. Legally, however, it has no control over the quality or even the numbers of those who elect to join engineering courses. The *Abitur* gives anybody the right to sign on for what he chooses, and that is that.

Nor, for the time being, does RWTH appear to have much interest in leavening the education of its engineers with a dash of the humanities (along the lines of the 20 per cent of time students at Massachusetts Institute of Technology are required to spend following courses in the humanities). As it happens, RWTH almost accidentally acquired a philosophy faculty in the heyday 1960s (together with a teachers' training college), and the senate will soon have to decide whether to respond to the financial crisis in Nordrhein-Westfalia by taking the knife to that cuckoo in the nest. The news that the Electrotechnische Institut in Zurich, with which Aachen has close links, is making a few experiments of this kind has unsettled hard-liners.

RWTH seems also to have been comparatively untouched by the wave of recent university reform. In 1968, it was slow to follow other West German *Technischen Hochschulen* in calling itself a university. On reflection, people are glad that they have kept the name first coined in 1870. And while there are student representatives on most committees, their proportions are such that they can be voted down when necessary. □

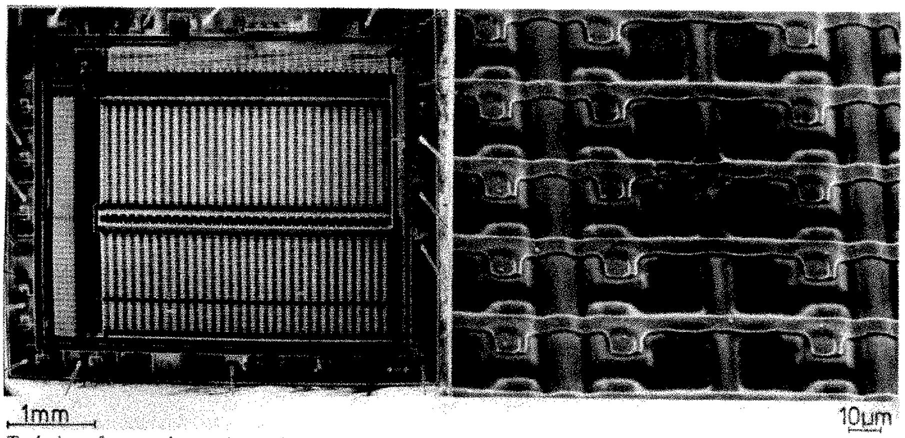
Where graduates make good

SIEMENS AG, the largest employer of scientific manpower in West Germany, is not so much a source of monthly pay cheques as a way of life. The company offers those who work for it attractive fringe benefits such as pensions, while it makes a point of recruiting all its senior managers from within its own ranks. Indeed, the company boasts that the members of its board of directors are almost all technical people who have worked their way up within the hierarchy — and that there are no lawyers, only money people, from outside. The result of all this is that recruits to Siemens who survive the first five years are likely to stay in what has become a remarkable example of a technical enterprise run by scientists and engineers.

So how does Siemens recruit its technical people? Of the 180,000 employees in Germany, more than 26,000 are technically qualified. In recent years, to make good losses and to sustain a growth of 2.5 per cent a year, the company has been recruiting about 2,000 graduates from the universities and the *Fachhochschulen*. What with the recession, this year's intake will be no more than 800, however.

In the German university tradition that students must fend for themselves, university careers offices do not exist. So a re-

0028-0836/82/210276-02$01.00



Technique for non-destructive testing of microcircuits developed by Dr H. Rehme at Siemens-Perlach. Left-hand panel represents secondary electron emission by an electron probe scanned over the surface. Dark areas are at positive voltage. The dark like across the lower half indicates the presence of a defect affecting a whole row of circuit elements and revealed in the electron micrograph (right-hand panel).

cruiting team from Siemens trudges round 20 universities and 60 *Fachhochschulen* twice each academic year. On balance, the company prefers to recruit engineers after the first degree (*Diploma*), simply on the grounds that it is easier to take on people in their twenties (the median age of university graduates in engineering is between 26 and 28) than in their thirties.

With the experience gathered since Werner von Siemens founded the company in the nineteenth century, the recruiting men now know where to go. Of the university recruits, 47 per cent are electrical engineers. Physicists make up 13.5 per cent of the university recruits, while the past few years have naturally seen a rapid growth of the recruitment of mathematicians and information scientists. Briefly, the company has a well-trodden selective path between the good engineering schools in West Germany, but also, interestingly, in Austria. Something like 10 per cent of its skilled labour force is from outside West Germany.

So how well is Siemens satisfied by the performance of the country's engineering schools? The specialization which is now common in the bigger and better schools is in some ways an embarrassment. The University of Karlsruhe, for example, produces graduates in electrical engineering under twenty different labels. Siemens approves of the objective — that a qualified engineer should know all there is to know in his or her chosen field — but wishes that it were possible also to help young engineers to be more flexible.

The consequences of the successive reforms of higher education are not welcomed. Siemens people shake their heads over the plight of some of the new universities such as Bremen, which have buildings and teaching staffs but which have failed to attract students of high quality in the numbers originally planned. They also agree with the widespread criticism of the decision that graduates of both the *Technische Hochschulen* and the *Fachhochschulen* should in future append to their names indistinguishable degrees — *Dipl. Ing.* for all. The result, they say, is that potential employers will not be able to rely on the uniformity of standards hitherto ensured by the various boards of engineering operating nationwide.

The company's chief concern, however, is that the attitudes of young people leaving engineering schools in recent years have been sharply different from those in the more distant past. People have been less competitive than they used to be, less eager to get on. They have also been corrupted by concern about the environment, a distaste for affluence and all that. But, so the argument goes, in the past year or so there has been a change. People are beginning to respond to the challenge presented by Japan. It's vital, said one company official, not just for West Germany but for the whole of the West that the Japanese should be challenged effectively at their own game of innovation.

The most striking feature of the company's use of its technical people is that close on 10,000 of its qualified staff work in research and development in laboratories scattered through the company's six operating groups but also in two central laboratories, one at Munich and one at Erlangen (not far away). The central laboratories are thus an important part of the cement that holds the six operating divisions together.

The scale of Siemens's investment in research and development should be sufficient to give it a sporting chance. For the company as a whole, 48 per cent of sales in 1979–80 consisted of products developed in the previous five years. Inevitably, the contributions of recently developed products are even higher in the company divisions responsible for components and information systems. Research and development cost more than DM3 million a year, exceeding 9 per cent of the company's sales (56 per cent of which are outside West Germany).

The central laboratory at Munich is housed in a complex of Lego-like buildings on the edge of the housing desert of the suburb of Perlach. Its managers say that they encourage proposals for research projects from members of the laboratory staff, and although most projects are oriented towards products that may ultimately be saleable, they insist that they can fairly be compared with the Bell Telephone Laboratories in New Jersey because of their willingness to take on basic research.

The laboratory also boasts of close links with the university system, pointing to its usual loss of one or two senior researchers to university chairs each year. And even the scientists and engineers who leave for jobs elsewhere are not an outright loss — they remain customers for Siemens products, it is hoped. □

No moving parts

ONE of the first instruments whose operation depends on the special theory of relativity is a spectacularly neat gyroscope built at the Siemens research laboratory at Munich-Perlach. Briefly, the device is a circular coil of single-mode glass-fibre around which radiation beams from a single semiconductor laser diode are sent in opposite directions. Rotation of the coil about its axis changes the velocities of the two opposing light beams, and thus the position of the interference fringes produced by recombining them.

The prototype instrument, built by Gerhard Schiffner (now at the University of Bochum), Benhardt Nottbeck and G. Schöner, consists of 2.88 kilometres of optical fibre wound on a 12.3 cm diameter former. To obtain an analogue signal proportional to the rate of angular rotation of the coil, the phase of the counter-travelling light beams is modulated by means of a short length of fibre whose length is modulated by means of a piezoelectric element. The signal produced by the recombined beams at a silicon photodiode is amplified and analysed by means of two lock-in amplifiers, one at the modulation frequency and one at twice that frequency. The ratio of the two signals can in principle provide a measure of rotation rate independent of power output or of the attenuation of the glass fibre.

The obvious rotation rate to measure is that of the Earth — some 10 degrees per hour for an axis pointing north–south at the latitude of Munich. Measurements of this rate suggest that the sensitivity of the prototype instrument is limited by noise to about 0.3 degrees per hour, while drift of the calibration scale may be about 0.5 degrees per hour. Temperature fluctuations (affecting the length of the optical fibre and its optical properties) and even fluctuations of the Earth's magnetic field may contribute to drift.

Siemens has little to say about the potential applications of the instrument, but the attractions of what may be a sensitive gyroscope with no moving parts are obvious. The immediate problem is to achieve miniaturization, using fibre optics elements as means of constructing the interferometer which is the essence of the instrument.

0028-0836/82/210277-01$01.00

Helping small businesses

SMALL business is big in West Germany, respected and much cherished. The best-known companies (Bayer, Siemens, Krupp and so on) are huge, and also big spenders on research and development, but everybody knows that the real strength of West German industry rests with the small and middle-sized companies, often too small to support research and development programmes of their own. What can the governments, federal and regional, do for them?

Quite a lot, it seems. West Germany offers a richer menu for encouraging research on a shoestring than any remotely comparable country. Given the way in which graduates of the engineering schools slide into the management of the small companies (and behind the wheel of the accompanying Mercedes), it is natural that they should first turn for help to the academic departments whence they came.

The volume of direct research contract business at the large engineering schools is, however, small. But the constitution of the universities as dependants of the Länder makes the connection between the universities and local industry a two-way street. A large but unknown proportion of the grant applications to DFG for engineering projects is stimulated by the needs of small business.

The federal and *Länder* governments also support a more formal mechanism for collaborative research — the Fraunhofer Society (Fraunhofer-Gesellschaft zur Förderung der Angewandten Forschung). This organization maintains some 27 research institutes in West Germany, with interests primarily in sold-state electronics (including defence applications), information processing and the construction and woodworking industries.

Contract research is now the most rapidly growing part of the Fraunhofer Society's work, apparently stimulated by the agreement struck between the federal and *Länder* governments in 1975 on the scale of their support (now running at about a third of the total budget of DM250 million).

In the past five years, the research and economic ministries of the federal government have gone even further and have introduced a number of schemes which amount to direct subsidies for research by small businesses. Small companies letting research contracts to some external research organization (which may be a public laboratory) can apply for grants to cover a third of the cost, but up to a fixed amount. And there is also a scheme of grants to cover the costs of approved in-house research, again within a ceiling.

At the same time, the national research centres are being encouraged to take on research projects for small companies — and given the unaccustomed financial pressures to which they are now exposed, seem anxious to respond.

The overwhelming response to the research ministry's offer, at the end of last year, of small grants to help with the development of products incorporating microprocessors is one striking illustration of the willingness of small companies to play their part in what has become a self-conscious crusade for the application of research. Having offered grants amounting to DM300 million over three years (up to a maximum of DM800,000), the research ministry has been swamped with 1,400 applications in the first few months.

The common explanation for this enthusiasm is nothing if not encouraging. What people say is that West German companies, and especially the smaller among them, have awakened to the consequences for them in the present hiatus in the economic miracle and in the challenge from Japan. One small metal-working manufacturer at this year's Hannover Fair put the point simply: he reckons he has at most three years in which to redesign all his products.

Research for sale

MOST stands at the Hannover Fair attract visitors either by means of something that moves or by means of a young woman (rarely a man) sitting at a keyboard hitched to a computer or the like. Dr Jürgen Schulte-Hillen, by contrast, puts himself on show. From morning to night, he sits at a small table seeking to interest potential clients in the services of his small company as an adviser on research and development. Half-way through last month's fair, he was well on the way to laryngitis.

Management consultancy is not, of course, new in West Germany, while organizations such as Battelle (whose Frankfurt establishment includes one of the largest laboratories in West Germany)

Schulte-Hillen predicts a vigorous response to the challenge from Japan

will not merely advise a company on its need for research but will carry out the work on a contract basis. Schulte-Hillen, however, reckons that he has found a unique niche in West German industry.

His small company, now ten years old and employing only 20 people, deals principally with small and medium-sized companies. Its strong suit is to provide advice on how, in a client's operations, research and development might lead to a new product or process. Schulte-Hillen does not himself carry out the work, but will instead arrange a contract between the client and some research establishment, supervising progress much as an architect will breathe down a builder's neck . In the process, the argument goes, the managements of small companies learn what research and development can accomplish — training in research management, he calls it.

So what (if anything) is wrong with the state of science in West Germany? Schulte-Hillen was for six years head of project planning at a national aerospace laboratory (Deutsche Forschungs und Versuchsanstalt für Luft und Raumfahrt, see page 266) before setting up on his own. He believes that the national laboratories and Max Planck institutes are "highly qualified" institutions with deservedly high reputations, but that they are not flexible enough and that they should have closer links with industry or the universities, as appropriate.

Schulte-Hillen also shares the view common at least among those who graduated before the 1970s that the university reforms have been "a great mistake". He regrets that they have created two classes of universities in engineering, Aachen at one end of the spectrum and Bremen (everybody's whipping-boy) at the other.

The Japanese challenge to West German industry is preoccupying but not depressing. The reason for the Japanese success, Schulte-Hillen believes, is that they have set out to make well-engineered products for a mass market. Except in a few fields, German industry, based as it is on medium-sized companies, is not naturally well placed to compete.

But, in Schulte-Hillen's view, the climate is changing. The past few years have been a shock for West German industry, and the West German automobile industry for one is now bent on becoming competitive again by ensuring good design, high-quality and low costs.

Another sign of the changing climate, Schulte-Hillen believes, is the response of West German industry to the federal government's announcement last year of a programme of small grants (up to DM 50,000) to small companies with development projects intended to use microelectronic devices in novel ways. The applications received since January this year apparently already exceed the DM 300 million originally set aside for the next five years — and one suspects that Schulte-Hillen has written many of them.

0028-0836/82/210278-01$01.00

Research labs galore

THE legend, according to one senior official of the federal government, is that those who planned the reconstruction of West German science after the war deliberately decided to take over the conspicuous institutions in what were then the obvious models for the administration of science. So now there are research councils (DFG) on the British pattern and national laboratories, as in the United States. The Max Planck Society, however, is special and indigenous.

Whatever the explanation, large laboratories are now an important part of the West German scene, with the establishments run directly by the ministry of research and technology most prominent among them. Altogether, these twelve establishments employ more than 16,000 people and cost the ministry close on DM1,700 million a year.

The laboratories are a mixed bag, including as they do the Cancer Research Centre at Heidelberg (see page 267) and a centre for biotechnology research at Braunschweig (which employs fewer than 300 people). There is a case for thinking that the Cancer Research Centre would fit more comfortably under the wing of some part of the federal government more directly concerned with medicine.

The biotechnology laboratory, on the other hand, originally founded to meet the needs of the fermentation industries, now has a distinguished group of molecular biologists working on such projects as the transfection of cells in culture by gene fragments including the gene of beta-interferon. The laboratory is likely to be one of the beneficiaries of the DM70 million that the ministry of research has to spend this year on various aspects of biotechnology. It remains to be seen how far and how fast the laboratory will move along the path towards commercial exploitation. (As it happens, the laboratory was founded in 1968 with a grant from the Volkswagen Foundation.)

For the rest, the national laboratories range from largely academic establishments including the particle physics laboratory DESY at Hamburg (see page 270) and the plasma physics laboratory at Munich (see page 271), which is also a Max Planck institute, to the nuclear establishments at Jülich and Karlsruhe and the sprawling aerospace research organization centred on Cologne. They are collectively represented in Bonn by an Association of Big Science Establishments (or AGF for Arbeitsgemeinschaft der Grossforschungseinrichtungen).

As elsewhere, AGF's West German members are having painfully to adjust to changed economic circumstances. As recently as 1980, the laboratories were hatching a scheme for employing 700 younger scientists in specially created posts — a temporary expedient to keep young people in research and the research ministry's equivalent of the Heisenberg programme run by DFG. Now, like other government establishments, the national laboratories are having to reduce their staffs by 7.5 per cent over the next five years. Even so, AGF hopes to maintain the validity of its boast that 350 graduate students from West German universities earn their doctorates at its member establishments.

The process of contraction should not be as difficult as some fear. Natural wastage by retirement, death and so on runs at about one per cent a year, while the rate of resignation by people leaving for other jobs is said to average 4 per cent a year. The question that naturally arises, however, is whether there should be some more radical change in the pattern of this substantial research effort.

Much, no doubt, will depend on the success of these laboratories in persuading the general run of West German industry that developments incidental to what would be called, in the United States, their missions are of commercial value. Beneath an elaborate green tent-like structure at the Hannover Fair this year, AGF orchestrated an impressive display by its members of technology that may have a wider use.

Thus the Karlsruhe laboratory now boasts of a piston-driven pump for liquid helium, the Jülich laboratory of novel ceramic heating units (offshoots of its work on the development of equipment for installation in high-temperature reactors) and the aerospace laboratory claims already to have had success overseas in licensing its designs for fuel-oil burners. (Their high efficiency depends on the way in which droplets of fuel oil are vaporized by combustion gases before being ignited.)

The two nuclear laboratories at Karlsruhe and Jülich are of particular interest because of their close involvement, over past years, in the West German development of civil nuclear reactors. (Other member laboratories of AGF have an interest in nuclear affairs, including the GSF laboratory in Neuherberg near Munich concerned with problems of the geological disposal of radioactive wastes, the Hahn-Meitner institute in Berlin with an interest in radiation chemistry and the GKSS centre at Geesthacht-Tesperhude in northern West Germany which is a monument to enthusiasm in the early 1960s for nuclear-propelled ships.)

Both establishments have long since outlived their original remits. The Jülich establishment, for example, was founded (in 1956) by the *Land* government as a nuclear research facility for the universities of Nordrhein-Westfalen but was converted into a national laboratory in 1968. It is still responsible for the development of the technology of high-temperature reactors, as Karlsruhe is responsible for fast-reactor technology. (The plan is that Karlsruhe will shed 70 of its 270 fast reactor people in five years or so, when the prototype fast reactor at Kalkar should at last be complete.)

The question never far beneath the surface is whether there will be reactor-development projects with which to continue when these come to an end. For the past two years, even the construction of well-established types of reactors has become a contentious political issue — and not merely in the safe SPD country of Hamburg but in CDU Bavaria as well. Although Chancellor Schmidt was able to avoid the two-year moratorium on civil nuclear construction demanded of him at his party's conference in Munich last month, plainly the political climate is one in which new proposals for major reactor development would not be easily accepted, while the electricity utilities, stung by the research ministry's success at persuading them to contribute extra funds to the cost of the fast reactor prototype, are unlikely to be pressing for further initiatives in this direction.

The present plan seems nevertheless to be that these laboratories should be kept alive on more or less the present scale, with the understanding that new tasks should be transferred to them whenever possible. The principle that it is easier to start a new laboratory than to close one down applies in West Germany perhaps even more rigorously than elsewhere, given the close interest (and the 10 per cent financial contribution) of the *Länder* governments.

AGF is not the source of funds for these laboratories but rather a coordinating mechanism for those national laboratories that happen to come under the wing of the research ministry. Other federal laboratories exist more inconspicuously, again but in partnership between the federal government and the *Land* concerned.

Among these, the geoscience and resources agency at Hannover has made an international reputation for itself with its overseas surveys, especially in developing countries (and has also just completed a thorough survey of ground-water resources and their contamination by pollution for the whole of West Germany). The Tübingen veterinary laboratory similarly has a distinguished academic as well as practical reputation. Conspicuous by its absence in West Germany is the string of medical research laboratories that would be found in most other places — a measure of the dominance of the medical faculties but perhaps also a pointer to the need that medical research should be less fragmented.

On the principle that anything the federal government can do the *Länder* can do independently, there is also a spattering of regional research laboratories on the map of West Germany. Veterinary and geological laboratories are common, but Hesse supports the Sigmund Freud institute of psychoanalysis and Lower Saxony plans to set up a microelectronics laboratory once it has chosen between Hannover and Braunschweig. □

0028-0836/82 210279-01$01.00

What prospects for improvement?

THIS brief survey of science in West Germany is necessarily incomplete. Only a small sample of the important national research laboratories rates is mentioned. Most Max Planck institutes have been overlooked. And much too little has been made of the great wealth of sound biology now flourishing in universities, much of it deliberately fostered by the grant-making agencies such as the Deutsche Forschungsgemeinschaft. It is hoped that these deficiencies can be made good in the months ahead. For what is happening in West Germany, both in laboratories and in the committee rooms that help to shape their work, deserves much closer attention from the rest of the scientific community.

Incomplete though they may be, the preceding pages nevertheless confirm first impressions. Both researchers and those who manage their affairs are unsure of themselves, concerned that the substantial commitment of West Germany to research and development may be yielding less than it might. Comparisons with the United States are constantly preoccupying and rarely confident, which is why the doings of the particle physics laboratory at Hamburg (see page 270) stand out. Here at least is one laboratory not depressed by pointless calculations about the award of Nobel prizes. (For what little worth they have, the figures show that there have been 16 Nobel awards to West German nationals since the Second World War, compared with 124 and 42 awards to American and British nationals respectively.) The important question is how such a state of grace can be brought about elsewhere.

One of the dangers in West German introspection about the quality of research and development is that the distinctive and valuable features of the system will be undermined. Thus for decades other European countries have envied the way West German *Technische Hochschulen* forge such close connections with particular industrial sectors that both partners benefit immensely, as does the West German economy as a whole. Whenever British politicians take up the gloomy question "Why is Britain not more productive?", there is certain to be somebody to remind them that it might be different if British universities chose to follow this West German practice. The report (on page 275) from Aachen shows this practice still flourishes. The snag, for those who look for proof of success in the international award of honours, is that research in the design of machine tools is neither one of the topics mentioned in Nobel's notorious will nor a productive source of references in the primary journals. And while there may be some who complain that the education offered by the hot-houses among engineering schools may be undesirably narrow, or needlessly demanding, the success with which the graduates of these schools are able to make their way in West German industry is surely a proof that the engineering schools and West German industry have found a productive pattern for their relationship.

Elsewhere, however, the university system is in a mess. The reforms of the past quarter of a century have had the benefit of making sure that tenured professors cannot behave as tyrants, and have helped also to give West German universities a sense that their corporate identity extends to all who work in them, but only at the cost of licensing an intolerable and demeaning bureaucracy. Is it any wonder that people seeking a career in research prefer to work elsewhere, at a Max Planck institute perhaps, or even in the United States? The most obvious difficulty is that ridding the system of this blight cannot be accomplished by the universities themselves, only by the twelve governments of West Germany.

Sooner or later, something will have to be done to allow the universities to match their intake of students in all fields to their capacity to teach them properly. The pretence that all universities are educationally and intellectually equal will have to be abandoned. The universities must be rid of the legal harassment by students and would-be students with which they are now plagued, and should be free to decide for themselves the composition of the internal committees that manage their affairs. Nobody pretends that such steps could be taken easily, for they would offend against cherished constitutional principles. But to fail to take them is to perpetuate bumbledom and inefficiency.

There is also a need to change the present balance between the support for basic research by the principal grant-making agencies. The most glaring need is for a closer link between the 6,000-odd researchers who work in the Max Planck institutes and the much greater number of professional scientists who work in universities. Although the Max Planck Society acknowledges the need that new institutes should be sited near universities, although many institutes have members who work on common problems with university colleagues, and although senior members of the institutes are eager (and welcome) teachers at their local universities, the plain truth is that the institutes attract many of the most able younger scientists — and then provide many of them with an environment in which they can securely become less productive. The resentment that has sprung up between the two groups of people is not good for the research enterprise.

There is a need, less obvious but just as urgent, for adjusting the present balance of support for research within the universities. One of the unusual features of the West German system is that the bulk of the cost of university research is met directly by the *Länder* (regional) governments, with the Deutsche Forschungsgemeinschaft playing a relatively minor role. (The official estimate of a ratio of four-to-one is probably an over-estimate, but is not easily verifiable.) At many universities, Heidelberg for example, tenured professors look to the general budget for research support, but more junior colleagues must take their chances with DFG. Could this be why people returning to West German universities after a promising start elsewhere are so often said to slump into inactivity? The doctrine of the unity of research and teaching is wholly admirable but there is a danger that much of the research support now channelled from the regional governments to university researchers finds its way into projects that would not survive the rigours of peer review. This may also explain why the de-deification of the old-style professors has not been matched by a corresponding sense of liberation among junior colleagues. Given the generality of the observation that, compared with their equivalents elsewhere, younger scientists in West Germany are less adventurous — less likely to be disastrously wrong or brilliantly right — a more selective way of offering larger dollops of research support seems necessary. The general anxiety about clinical research argues in the same direction.

The problems peculiar to West Germany appear principally to affect basic research. There are also the familiar problems of what to do with research establishments that have lasted longer than the problems they were intended to solve, arguments about research strategy (how much to spend on solar energy development?) and anxieties about political interference (should a new laboratory go in one place rather than another?). West Germany is not worse afflicted than other places, while the scale of its research enterprise, and the skill of its practitioners, imply immense advantages. So has not the time come for the research community to insist that what it knows to be wrong should be put right? The paradox of West Germany is that the elaborate hierarchy of policy-making bodies that has evolved since the Second World War is somehow incompetent to challenge prevailing ideologies, even those that lead directly to the absurdities now manifest in the university system.

So the hunt goes on for other explanations of why West German science is not the glittering enterprise it should now be. It is like a heavy smoker with a chest infection seeking a psychosomatic explanation for his complaint. Although West German science may to some extent still be made tentative by the awful experience of what is euphemistically called "the period 1933-45", and by the sheer newness of the federal republic, what matters now is that each year increases the distance from that time — and brings a new cohort of productive people to adulthood.

0028-0836/82/210280-01$01.00

NEWS AND VIEWS

Crust to crust, basalt to basalt

from Peter J. Smith

As the author of Ecclesiastes might have put it, of the making of many geochemical hypotheses for the workings of the Earth's interior there is no end. Or so it seems. Those who don't pay the closest possible attention to the flow of geochemical thinking — and many of those who do, for that matter — can hardly be anything but bewildered at the sheer quantity of it all and the seemingly infinite number of paradoxes and inconsistencies it represents. New sites are continually being investigated, only to provide solutions to some problems whilst revealing fresh anomalies. New methods are continually being introduced, only to fulfil some of their promise but present additional difficulties. New hypotheses are continually being proposed, only to reconcile some data at the expense of mutually exclusive hypotheses that reconcile other data. To the outsider, ever more research must look suspiciously like the source of ever more confusion, offering the reminder that the Earth's interior only looked remarkably simple when next to nothing was known about it. The much-needed global model that succeeds in reconciling the mass of conflicting observational evidence seems as far away as ever.

The new model put forward by Hofmann and White[1] hardly fills that all-embracing role, but, if valid, it should go a long way towards explaining some of the apparent contradictions that have recently come to the fore as a result of the increasing attention being paid to oceanic island basalts (OIB). Not least among the inconsistencies is the fact that, whereas OIB are 'enriched' (that is, with respect to 'primitive', bulk-earth, chondritic material) in incompatible elements such as K, Rb, U and Th and the light rare earth elements (LREE), Nd isotopic data seem to suggest that the source of most OIB is 'depleted' in LREE. For some time now, the differences between OIB and mid-ocean ridge basalts (MORB) have been interpreted as indicating distinct sources in the mantle; and under those circumstances it was possible to hold (not that everyone did) that the OIB were derived from primitive material (such as the lower mantle, if primitive it be) and the MORB came from a relatively depleted source (for example, the asthenosphere, if depleted it be). But the Nd data would seem to rule that out.

So how can LREE-enriched OIB be, or appear to be, derived from an LREE-depleted mantle source? Wasserburg and DePaolo[2] have suggested that perhaps the Nd isotopic data only indicate a non-primitive source for OIB because the initially primitive material gets contaminated with depleted MORB source material in the upper mantle. But that appears to conflict with Pb isotope and other evidence. Menzies and Murthy[3], among others, have proposed that metasomatism enriches the OIB source before melting. But there appears to be no evidence for that, at least from the source material itself. Even earlier, Gast[4] had attempted to account for the enrichment of OIB by supposing that primitive OIB source-material undergoes less melting than MORB source-material. But the degree of melting this requires is probably far lower than most petrologists would countenance.

Hofmann and White cut right through these proposals and objections. They first state that the OIB source cannot be primitive — a conviction based on the "rather conclusive" evidence that (1) no primitive source could give rise to enriched OIB without an unacceptably low degree of melting, and (2) a primitive source is inconsistent with the isotopic data (Rb, Nd, U, Th and so on). Next, they express their faith in Morgan's mantle plume hypothesis[5], because they find "his arguments as persuasive today as they were ten years ago and because . . . no viable alternative model has been proposed that would explain the kinematic and geochemical peculiarities of 'hot-spot' volcanism". They then look round for a possible source of the non-primitive material that forms the mantle plumes, and come up with the bold suggestion that it is ancient, subducted crust.

As might be expected, this is not the first time that the fate of subducted crust has been considered. Ringwood[6], for example, suggested that such crust might sink to the base of the mantle and thus be lost permanently to the convective system. Tatsumato and Knight[7] and, more recently, Anderson[8] have proposed that subducted crust is recycled to form new oceanic crust, including OIB. What Hofmann and White have done, in effect, is to combine components from both these views into the proposal that subducted crust sinks to the bottom of the mantle whence it ultimately rises again in the physical form of mantle plumes.

As new, initially basaltic, oceanic crust moves away from ridges it is subjected to hydrothermal alteration, undergoes low-temperature alteration and accumulates sediments; and as it subducts it experiences metamorphism and partial melting. Even so, its bulk composition is still likely to be more or less basaltic, and its density will almost certainly be higher than that of upper mantle material and probably higher than that of lower mantle material too. It therefore sinks until it reaches a level of density compensation (presumed here to be as deep as the core–mantle boundary), shedding the more refractory part of the lithosphere as it does so. Once at its new site, where it builds up to thicknesses of 100 km or more, it begins to heat up, partly because of radioactive heat production within it and partly through heat transfer from the core. As the temperature rises the layer becomes unstable, and diapirs (plumes) ascend into the upper mantle where the material melts, rises to the surface and forms oceanic islands or continental hotspots as appropriate.

Hofmann and White go to some lengths to demonstrate that such behaviour is thermally and mechanically plausible. But something more than that emerges, for the (admittedly crude and assumption-ridden) calculations also give a hint of the time scale involved. The time taken for the once-crustal layer at the base of the mantle to develop instability seems to be as much as

Peter J. Smith is a Reader in the Department of Earth Sciences, The Open University, Milton Keynes MK7 6AA, UK and editor of Open Earth.

0028-0835/82/210281-02$01.00

one to two billion years, implying that the material now rising as mantle plumes was oceanic crust as long ago as the Proterozoic. In this connection, it is interesting to note that apparent mantle isochrons for recent OIB commonly yield ages of about 1.8 (Pb) and 1.6 (Rb/Sr) billion years. Moreover, Morgan also suggested that plumes may actually initiate seafloor spreading/upper mantle convection by breaking up plates and thus allowing the passive upwelling of mantle material. Putting all this together, it is tempting to envisage an Earth in which periods of plume-triggered mantle convection alternate with intervals of mantle stagnancy during which the stored, subducted crust gradually heats up towards instability.

The first of the geochemical conclusions to be drawn from all this is that, by definition, OIB source-material cannot be primitive. On the other hand, the exact bulk composition of the material is difficult to specify. As the crust subducts it may or may not take sediment with it, it will be metamorphosed to an uncertain extent, and dehydration and partial melting will remove an unknown quantity of incompatible elements. At some point it will then be converted to eclogite and the less dense, depleted part of the lithosphere will separate and return to the MORB reservoir (presumably the asthenosphere). It seems unlikely, however, that these processes will be anything like thorough enough to remove all the incompatible elements — a conclusion for which there is some support from a consideration of what happens to K during subduction and magma production in island arcs. In other words, it would seem that when the subducted crust reaches the lower mantle it is still enriched.

The problem of having to explain how enriched OIB are derived from a depleted source is thus removed; the source is not depleted. What, then, of the Nd isotopic data, which seem to demonstrate otherwise? At this point it is necessary to bear in mind that, according to the Hofmann–White hypothesis, modern OIB are derived from Proterozoic crust — that is, Proterozoic MORB. Studies of mantle-derived basalts have shown that Nd/Sm ratios in large parts of the mantle have become progressively depleted. Thus Proterozoic MORB were less depleted than are modern MORB; and, as a result, modern OIB would be expected to have variable but generally lower $^{143}Nd/^{144}Nd$ ratios than modern MORB — which they do. Moreover, a MORB source would be able to produce OIB with 'normal' degrees of melting — 10 to 30 per cent for tholeiites.

As for the major elements, Hofmann and White point out that the idea of an OIB source of an essentially basaltic composition is "not dear to petrologists", the prevailing view being that basalts are produced by the partial fusion of peridotite. There are indeed problems in envisaging the generation of basalt from eclogite; but they may not be insuperable in this case insofar as the Hofmann–White model does not insist upon a source of pure eclogite. The subducting and sinking crust may not shed quite all of its associated refractory lithosphere. The incorporation of, say, 10–20 per cent of mantle peridotite would be sufficient to overcome the geochemical objections to an eclogite source (such as the fact that most OIB are olivine saturated). Moreover, there is bound to be some interaction between the ascending mantle plumes and the upper mantle peridotite through which they pass. None of this would change significantly the density, incompatible element composition and heat production in the rising plume material but it would easily provide sufficient olivine and pyroxene to bring the major element (and compatible element) composition into line with observation. To this extent, therefore, major element abundances do not impose strong constraints on the new model.

It is a measure of the disarray in which geochemistry finds itself that almost any sentence of substance in any article on the subject can, and is likely to, attract perfectly legitimate objections. Hofmann and White's work is unlikely to be an exception. Indeed, they admit that their model is an extreme one, "a broad target for criticism", a bold idea that may have to be modified if some of the details fail to correspond with future observations. Few people, however, would deny the importance of bringing the question of OIB–MORB differences to the fore. Many, if not most, geochemical models of the mantle have hitherto ignored OIB, largely on the grounds that they are volumetrically small in comparison with MORB. What Hofmann and White have done is show that relative volume may be no measure of scientific significance. □

1. Hofmann & White *Earth planet. Sci. Lett.* **57**, 421 (1982).
2. Wasserburg & DePaolo *Proc. natn. Acad. Sci. U.S.A.* **76**, 3594 (1979).
3. Menzies & Murthy *Earth planet. Sci. Lett.* **46**, 323 (1980).
4. Gast *Geochim. cosmochim. Acta* **32**, 1057 (1968).
5. Morgan *Nature* **230**, 42 (1971).
6. Ringwood *Composition and Petrology of the Earth's Mantle* (McGraw-Hill, New York, 1975).
7. Tatsumato & Knight *Geochem. J.* **3**, 53 (1969).
8. Anderson *Geophys. Res. Lett.* **6**, 433 (1979).

The gaps in the fossil record

from David E. Schindel

THE fossil record has been both a burden and a blessing to evolutionary biologists — a blessing because the stratigraphical order in which fossils are arrayed provides broad proof of change with time, but a burden in that the gradual morphological transitions between presumed ancestors and descendants, anticipated by most biologists, are missing. Darwin and most of his followers have resolved this dilemma by accepting the blessing and dismissing the burden as an unfortunate but inherent limitation in the fossil record resulting from varying rates of sediment deposition.

In their much cited paper, Eldredge and Gould (ref. 1; see also ref. 2) re-examined the discrepancy between palaeontological patterns and biological processes. They argued that despite the imperfections in the fossil record, the pattern of abrupt appearances of new species might be a faithful rendition of speciation events, if viewed on a geological time scale. They contrasted two models, phyletic gradualism and punctuated equilibria. In the former, new species arise through the steady accumulation of microevolutionary change over significant spans of geological time. The discrepancies between generation-to-generation change and the observed palaeontological pattern would therefore be real, since most species appear abruptly in the fossil record. In the punctuational view, species persist more or less unchanged throughout their duration but infrequently and abruptly give rise to new forms. The long spans of morphological stasis within species would thus be a literal record, not an artefact of an imperfect record. The discrepancy would be more perceived than real, in this view.

Over the past ten years, microstratigraphical studies have appeared which document fine-scale patterns of change. Using the finest scale of sampling possible, investigators have attempted to test the reality of the discrepancy between microevolutionary changes (visible during a human lifetime) and the palaeontological patterns by approaching the time scale on which biological processes operate. These field studies have faced two serious challenges in the past year. First, the scale on which evolutionary patterns are documented is critical, because "one man's punctuated equilibrium may be another's evolutionary gradualism"[3], and "stepwise changes can be made to appear gradual if a coarse enough time scale is selected"[4]. The second and more serious challenge states that the stratigraphical record, as a whole, is so incomplete that fossil patterns are meaningless artefacts of episodic sedimentation[5]. This article provides a response to these challenges by outlining procedures for evaluating the level of time resolution obtained in

David E. Schindel is a Curator of Invertebrate Fossils in the Peabody Museum of Natural History and an Assistant Professor of Geology at Yale University, PO Box 6666, New Haven, Connecticut 06511.

0028-0836/82/210282-03$01.00

microstratigraphical studies, and the quality of the patterns they document.

Three criteria are necessary for evaluating temporal resolution in the fossil record. *Temporal scope* is the total time span over which the pattern is documented. *Microstratigraphical acuity* is the time span represented by each individual fossil sample. *Stratigraphical completeness* is the proportion of the total time span (temporal scope) that is represented by strata.

Temporal scope is estimated with the common tools of stratigraphy, ultimately being established by radiometric dating techniques. Acuity and completeness estimates are based on short-term sedimentation rates, which document the rate at which sediments accumulate without gaps. Sedimentation rates based on longer-term observations include periods of non-deposition and erosion, and are therefore lower than short-term rates in similar settings. Brief periods of non-deposition or erosion are undoubtedly common, but are less likely to be described in the literature than are brief periods of active accumulation. The pattern of lower rates of sediment accumulation with longer spans of observation time has been documented in two recent compilations[6,7].

If individual fossil samples are taken from stratigraphical intervals free of evident sedimentological breaks (such as scoured surfaces, hardgrounds, bedding planes, attritional shell and bone beds), then dividing the sample thickness by a short-term sedimentation rate typical for that depositional setting (over, say, 100 yr) gives an estimate of acuity that assumes the interval of strata accumulated without interruptions greater than 100 yr (see refs 6 and 8 for further explanation). Completeness is estimated by first dividing the thickness of the entire strata by the appropriate short-term sedimentation rate, and the time span so derived is then divided by the temporal scope of the entire sequence. This completeness ratio represents the portion of the temporal scope represented by active accumulation of sediment, that is, if the entire sequence accumulated without interruptions greater than the time span of observation for the short-term rate used in the calculation. Sadler[7] provides a detailed and elegant treatment of this procedure.

The table presents data on the stratigraphical thickness and temporal scope of seven published microstratigraphical studies[9-15], along with my estimates for the acuity and completeness inherent in each. Three points emerge from this review of published studies. First, there is a trade-off between scope, acuity and completeness: a study can provide fine sampling resolution, encompass long spans of geological time, or contain a complete record of the time span, but not all three. Second, some microstratigraphical sampling schemes can provide acuity that approaches the time scale of biological processes (decades to centuries), if individual samples are collected with an eye to avoiding sedimentological breaks[6]. (Of course, extra caution is necessary in strata that show signs of bioturbation or resuspension, which act to obscure evidence of breaks.) Third, it is most difficult to evaluate evolutionary patterns in studies with high acuity (minimum of time per sample) and low completeness, unless the investigators document where the breaks occur in the sequence, and attempt to sample between the gaps. In so doing, a procedural trade-off is made by decreasing the scope of the sampled interval, within which acuity and completeness can remain high. This seems to be the case for the study by Williamson[15], and the within-cycle sampling by Schindel[14]. The patterns distilled from these relatively short, complete sequences can be read directly, and can be calibrated in years using short-term sedimentation rates. Without specific knowledge of where the breaks are located it is difficult to judge whether the other documented evolutionary patterns are gap-filled (and therefore may be artefacts of varying sediment input), or if they are relatively gap free, with the unrepresented time concentrated above and below the critical sampled portions of the sequences.

The debate over tempo and mode should proceed beyond the stage of arguing, in terms of generations, how fast is fast, and how slow is slow. It seems possible to estimate the quality of local stratigraphical sequences and the level of sampling acuity achieved in each. The gaps in the stratigraphical record are neither scattered so randomly nor hidden so well as to render the entire fossil record useless. I would therefore replace van Andel's[5] counsel of despair with a call for caution in sampling, and greater awareness among palaeontologists to the progress being made in genetic stratigraphy and facies analysis.

One other point emerges from this treatment of the gaps in the record. Most sedimentological breaks reflect changes in physical environmental conditions (water depth and current regime, sediment influx, carbonate productivity). Certainly, when one considers the large proportion of unrepresented time in the studies reviewed here, it seems very likely that changes in habitat conditions are scattered throughout stratigraphical sections. For this reason alone, one might question the wisdom of extrapolating the generation-to-generation processes witnessed by biologists (primarily during periods of stable conditions) up to geological time spans.

Contrary to a recently expressed[4] and commonly held opinion, the debate between the punctuated and gradual viewpoints is not an argument over rates of change, but an argument over the persistence of rates. Many biologists assume that the generation-to-generation microevolutionary processes they witness are typical short segments of trends that persist through geological time. Using this assumption, morphological stasis can be explained as the result of persistent stabilizing selection, and gradual trends in size or shape would result from sustained directional selection. But the present treatment of the gaps in the fossil record calls into question this assumption of persistent rates, stable environmental conditions and sustained selective regimes. Stasis may result if directional trends are short-lived and frequently reversed. Long-term trends may be processions of morphologically distinct and separate populations, rather

Temporal resolution achieved in seven published microstratigraphical studies, evaluated with three criteria defined in the text

Subject (ref.)	Stratigraphical thickness(m)	Temporal scope(Myr)	Sample thickness(m)	Microstratigraphical acuity(yr)	Time represented (yr)	%Stratigraphical completeness
Eocene mammals (9)	520	3.5	10	4,000	980,000	28
Permian forams (10)	200	12	≪1	600	500,000	4
Neogene radiolaria (11)	5	1.9	≪0.01	40	35,000	2
Neogene forams (12)	200	8.3	0.02	80	1,700,000	23
Jurassic ammonites (13)	14	1 +	≪0.01 (e)	40 (e)	28,000	3
Pennsylvanian snails (14)	1,300 (a)	27	≪10	5,000	9,300,000	34
	10 (b)	(5kyr?)	≪0.1	50	(5,000?)	(100?)
Plio-Pleistocene molluscs (15)	265 (c)	0.4	0.25	170	180,000	45
	110 (d)	0.1	0.25	170	73,000	73

Stratigraphical thickness, temporal scope and sample thickness are taken directly from original studies. Microstratigraphical acuity was calculated by doubling each sample thickness to correct for compaction, then dividing by a short-term accumulation rate for uncompacted sediment in that environment estimated from compilation of rates by Schindel[6]; these rates were based arbitrarily on 100 yr periods of observation, and therefore assumed that samples contained gaps no greater than 100 yr. Time represented was estimated by doubling the stratigraphical thickness to compensate for compaction, and dividing by the sedimentation rate appropriate for that environment and the level of sampling acuity, giving the time represented by each entire sequence if it contained gaps no greater than the time span of a sample. Completeness is the percentage of the temporal scope represented by active sediment accumulation at this level of sampling resolution. Note trade-off between scope, acuity and completeness.

(a) Among-cycle study spanning Middle and Upper Pennsylvanian time.

(b) Within-cycle study of a single mudstone interval in the Middle Pennsylvanian. The interval is entirely within a single fossil zone, so temporal scope and completeness cannot be estimated by independent evidence, indicated by (?).

(c) Entire Koobi Fora Formation.

(d) Lower Member, Koobi Fora Formation.

(e) Many samples were collected from bedding planes, and are probably time-averaged attritional assemblages. Estimates presented assume 1cm sample thickness within which breaks are not found.

than gradual modifications within a single standing population.

I do not question the reality of short-term directional changes, but I point out the evidence that abounds in the fossil record for habitat shifts, local extinctions and the general lack of persistence in physical conditions. One might assume that changes in physical habitat conditions occur slowly and imperceptibly, allowing any particular biota to migrate passively within geographical limits of the moving habitat. But this view would be an intuitive notion, not a proven point. The effect of habitat shifts on populations over geological time remains an open question. Microevolutionary processes probably operate within the constraints imposed by higher-order processes, which define the distribution and duration of habitable area available to organisms. The gaps in the record suggest that these processes disengage the short-term fate of individuals from the longer-term fate of populations, species and higher taxa. ☐

1. Eldredge, N. & Gould, S.J. in *Models in Paleobiology* (ed. Schopf, T.J.M.) 82 (Freeman, Cooper, San Francisco, 1972).
2. Vrba, E.S. *S. Afr. J. Sci.* **76**, 61 (1980).
3. Jones, J.S. *Nature* **293**, 427 (1981).
4. Ginzburg, L.R. *Paleobiology* **7**, 426 (1981).
5. van Andel, T.H. *Nature* **294**, 397 (1981).
6. Schindel, D.E. *Paleobiology* **6**, 408 (1980).
7. Sadler, P.M. *J. Geol.* **89**, 569 (1981).
8. Schindel, D.E. (in preparation).
9. Gingerich, P.D. *Am. J. Sci.* **276**, 1 (1976).
10. Ozawa, T. *Mem. Fac. Sci. Kyushu Univ., Ser. D. Geol.* **23**, 117 (1975).
11. Kellog, D.E. *Paleobiology* **1**, 150 (1975).
12. Malmgren, B.A. & Kennett, J.P. *Paleobiology* **7**, 230 (1981).
13. Raup, D.M. & Crick, R.E. *Paleobiology* **7**, 200 (1981).
14. Schindel, D.E. *Bull. geol. Soc. Am.* **93**, 400 (1982).
15. Williamson, P.G. *Nature* **293**, 437 (1981).
16. Wiens, J.A. *Am. Sci.* **65**, 590 (1977).
17. Boag, P.T. & Grant, P.R. *Science* **214**, 82 (1981).

The age of galaxies

from M. G. Edmunds

ALTHOUGH it has become apparent in recent years that there are quite a few observational constraints on cosmological models, it remains true that facts in cosmology are rare. Any new information must be greeted with delight, and two recent investigations of time and distance scales in the Universe are particularly welcome.

The observational evidence for the evolution of galaxies over cosmic time dates back to the interpretation of radio source counts. Subsequently the so-called 'luminosity–volume' test was applied with considerable success to suitable samples of quasars, and it was concluded (well reviewed by J.V. Wall in *Phil. Trans. R. Soc.* **A296**, 367; 1980) that very active objects like quasars and radio galaxies were far more numerous in the past. It is not suggested that these sources have now completely disappeared, but rather that many galaxies which are comparatively quiescent at the present time must have been much more active at earlier times. The general indications are that the effective number of certain kinds of sources was greater by a factor of perhaps 1,000 when we look back to redshifts between 1 and 3 (corresponding to about 1/3 to 1/5 of the present age of the Universe, assuming a standard expansion model with fairly low overall density of matter).

It has been realized for several years that back beyond a redshift of 3 there may well be no further increase in source numbers, but the observational problems of carrying out a survey for objects that far away are severe. A significant step backwards (if that is not an impolite way of describing the work!) has been made by P.S. Osmer (*Astrophys. J.* **253**, 28; 1982). He has searched systematically for quasars with redshifts greater than 3.5 (that is, looking back beyond 1/5 of the age of the Universe) in an area of 5 square degrees in the sky by using a special spectrum disperser in front of photographic plates. The disperser, a grating prism, stretches out the images of stars and galaxies into small spectra of just sufficient resolution to allow possible quasars to be identified by their strong emission lines. These faint quasar candidates were then examined at higher resolution with a conventional telescope/spectrograph combination. No quasars were found in the search area between redshifts 3.7 and 4.7 (the limit of the technique). Of course the survey has a faintness limit, but previous work in the redshift interval 2.5–3.5 suggests that at least nine quasars should have been discovered even if the space density of objects were constant beyond redshift 3.5. The fact that none was found implies that the density must decrease significantly. The lack of quasars in this particular area of sky could be attributed to a chance fluctuation in spatial distribution, but if we assume a Poisson distribution on the sky, then for an expected number of nine objects in the area, the variance is nine and the detection of none is a very highly significant result!

The importance of the definite demonstration of a decrease in space density lies in the implication that violent activity in galaxies starts at a time corresponding to around redshift 3.5. Since the activity requires a galaxy we can conclude that galaxies must exist at that time. Furthermore, as activity is thought to begin fairly rapidly after the actual formation of the galaxy as a discrete entity of stars and gas, it can be argued that we are seeing back to the epoch (about 1/5 of the age of the Universe) at which galaxies form.

Despite his careful area-limited search, Osmer must have been disappointed that he did not beat the record for the largest redshift discovered, set at 3.53 about a decade ago. It is perhaps a little ironic that very soon after Osmer's paper was published, a press release appeared from workers in Australia (B. Peterson, A. Savage, D. Jauncy and A. Wright), reporting that optical spectroscopy of the radio source PKS2000 - 330 shows that it is a quasar with redshift 3.78. This in no way vitiates Osmer's result, since it is not at all surprising that there are some quasars beyond redshift 3.5 — but there are not very many of them.

How long ago, in real terms, was this epoch of quasar and, by inference, galaxy formation? The expansion time scale of the Universe has been a controversial issue for many years, the uncertainty arising from the great difficulties of accurately measuring distance in astronomy. The conventional chain of reasoning is a long and indirect one, passing from the parallax measurements of the nearest stars, out to various size and brightness indicators in galaxies of known redshift. The number determined observationally is the Hubble constant, H_0, and the two major determinations of its value are either 50 km s^{-1} Mpc^{-1} or 100 km s^{-1} Mpc^{-1}, depending on whether the Universe is viewed from California or Texas. The pragmatic astronomer might be tempted to take a mean value between the two, and some support for this approach comes from a recent determination of the distance to supernova explosions in distant galaxies by W.D. Arnett (*Astrophys. J.* **254**, 1; 1982). The extreme brightness of supernovae makes them ideal probes of distance and although distance estimates have been attempted before, using kinematic models of their expansion, this is the first time that a dynamical model for type I supernovae has been used. The model is admittedly very simplified, and aims at estimating the intrinsic brightness of the supernova on the basis that the luminous energy is supplied by radioactive decay of ^{56}Ni, freshly synthesized in the explosion. The distance can then be inferred from the supernova's apparent brightness. From a set of supernovae in several clusters of galaxies, a value for H_0 of 70±10 is deduced, which lies nicely between the disputed values. The implied age of the Universe is then about thirteen billion years, again assuming a fairly low mass density. Arnett stresses that his value for H_0 should not yet be taken too literally, since there are many refinements to his simple supernova model that need investigation. Nevertheless, the age of galaxies would then be of the order of ten to eleven billion years, an acceptable time scale compared with present estimates of the ages of some star clusters in the Galaxy,

M.G. Edmunds is in the Department of Applied Mathematics and Astronomy, University College, Cardiff CF1 1XL.

0028-0836/82/210284-02$01.00

and the age estimates based on comparative abundances of radioactive nuclei. This would at least avoid the problems created by the higher value of H_0 of 100 km s^{-1} Mpc^{-1}, which suggests a Universe age of nine billion years compared with star cluster ages which are estimated to be in excess of ten billion years. The lower value of 50 km s^{-1} Mpc^{-1} cannot be ruled out, although the age of galaxies (following Osmer's result) would have to be at least 14.5 billion years. □

150th anniversary of Gauss's first absolute magnetic measurement

from S.R.C. Malin and D.R. Barraclough

On 26 May 1832, the horizontal component of the Earth's magnetic field at Göttingen was 1.7820 $mg^{1/2} mm^{-1/2} s^{-1}$ (= 17,820 nT). Not particularly remarkable, you might think, except that this was the first ever absolute measurement of the geomagnetic field and its attainment had required much of the armoury of one of the world's greatest mathematicians.

Carl Friedrich Gauss (1777–1855) was a gifted child. His early mathematical feats led his father grudgingly to agree with his teachers that the boy should receive higher education rather than learn a trade, but Gauss was also an exceptional linguist and might equally easily have chosen to study philology. It was the satisfaction he derived in 1796 from demonstrating the construction of a regular 17-sided polygon with ruler and compasses that finally decided him on a career in mathemtics. In his doctoral thesis, he proved the 'fundamental theorem of algebra' (the first of his four independent proofs of this theorem), and most of his work at this time was similarly concerned with pure mathematics. Although this continued to be his main interest, he gradually branched out into many other subjects, including applied mathematics, astronomy, geodesy, optics, telegraphy and geomagnetism.

Soon after 1830, Gauss became interested in the adoption of a universal system of units for all physical quantities, and came to the remarkable conclusion that even magnetic intensity could be measured in terms of mass, length and time. With the practical assistance of Wilhelm Weber, professor of physics in the University of Göttingen, he set about devising an experiment to make this measurement. He explains some of his ideas in a letter to the astronomer Olbers written on 18 February 1832.

"I occupy myself now with the Earth's magnetism, particularly with an *absolute* determination of its intensity. Friend Weber conducts the experiments on my instructions. As, for example, a clear concept of velocity can be given only through statements on time *and* space, so in my opinion, the complete determination of the intensity of the Earth's magnetism requires to specify (1) a weight = p, (2) a length = r, and then the Earth's magnetism can be expressed by $\sqrt{p/r}$."

The experiment came to fruition three months later, after Gauss had slightly revised his dimensional arguments. The apparatus required is very simple — two bar magnets (A and B), one single-filament suspension in which either magnet can be fitted, a ruler, a clock and two weights. The method is also simple, though a number of refinements are required if its full accuracy is to be realized. The first part is the vibration experiment in which magnet A is suspended horizontally, and its period of oscillation either side of magnetic north measured. The period T is related to the magnet's moment of inertia, I, its magnetic moment, M, and the horizontal intensity of the geomagnetic field, H, by

$$T = 2\pi\sqrt{\frac{I}{M\,H}}$$

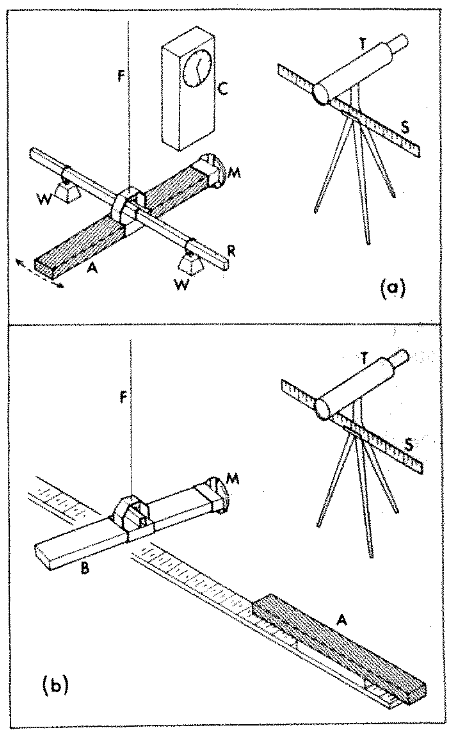
Thus, if I is known, we can infer $M\,H$. Gauss deduced I by adding known increments to it in the form of weights at known distances from the suspension, and noting how this affected T. In calculating I from the several different values of T and ΔI, Gauss used a powerful technique which he had developed himself, the method of least squares. He also took care to adjust T to the value it would attain with vibrations of infinitesimal amplitude, as the equation is strictly valid only for that case.

The second part is the deflection experiment in which the ratio M/H is determined by suspending magnet B and noting its angular deflection from magnetic north when magnet A is placed at a known distance to the east or west. This requires a knowledge of the way in which the magnetic intensity due to a bar magnet falls off with distance. Gauss deduced this from first principles, showing not only that the intensity due to a dipole depends on the inverse cube of the distance, but also that one additional term is both necessary and sufficient to allow for the finite length of the bar magnet. He determined the parameters by measuring the deflection with magnet A at different distances, again solving by the method of least squares.

From a knowledge of $M\,H$ and M/H, both M and H can readily be deduced.

Because of secular change, it is difficult to assess the accuracy of Gauss's original observation but, from internal consistency and backward extrapolation of the trend from later observations, it seems likely that it was within one per cent of the true value. With only minor modifications, Gauss's method continued to be the standard way of measuring H at magnetic observatories and in the field until the 1920s, and it is still in use at some observatories. Nowadays, absolute magnetic intensity measurements are usually made with a proton magnetometer. This gives the answer to an accuracy of about 1 part in 10^6 in a few seconds.

This was not Gauss's only contribution to geomagnetism. He later applied yet another of his inventions, the method of spherical harmonic analysis, to the Earth's magnetic field to show that nearly all of it came from inside the Earth. He was also, with Humboldt, largely responsible for the 'Gottingen Magnetic Union', an international project for the simultaneous observation of the magnetic field on four selected term-days each year at a number of widespread sites. This scheme developed into the worldwide network of magnetic observatories that exists today. □



Gauss's experiment. (a) The vibration experiment. The magnet, A, suspended by a silk thread, F, oscillates in a horizontal plane. Its moment of inertia can be varied by hanging weights, W, from the rod, R. The period of oscillation is obtained by observing the reflection of the scale, S, in mirror M, through the telescope, T, and timing an integral number of swings with the clock, C. (b) The deflection experiment. Magnet B is suspended and is deflected from magnetic north by placing magnet A at a known distance. The deflection is measured by observing the reflection of the scale, S, in mirror M, through the telescope, T.

S.R.C. Malin and D.R. Barraclough are in the Institute of Geological Sciences, Murchison House, West Mains Road, Edinburgh EH9 3LA.

0028-0836/82/210284-01$01.00

Cause and treatment of atherosclerosis

from C. T. Dollery

The most common cause of death in developed countries is necrosis of heart muscle (myocardial infarction) resulting from obstruction of the arterial blood supply. The affected artery is usually narrowed by a plaque of atheroma. Two main hypotheses have been proposed for the origin of the atheroma, and although they are not mutually exclusive, as became clear at a recent symposium*, each tends to have its particular supporters.

The first hypothesis holds that plasma lipids diffuse into and are deposited in the arterial wall, and receives support from epidemiological evidence that relates levels of blood lipids in younger people to incidence of atheroma. The second hypothesis is that focal damage to the endothelium leads to deposition of platelets from the blood, release of platelet-derived growth factor and stimulation of smooth muscle migration and hypertrophy. The laminated appearance of many atheromatous deposits, suggesting successive deposition of thrombus, and the presence of platelet antigens in the plaques provide supporting evidence.

In plasma, lipid particles consist of a core of neutral lipid surrounded by protein (the apolipoprotein). A.M. Gotto (Baylor College of Medicine) likened the apolipoprotein to a detergent that solubilizes the lipid and put forward the 'iconoclastic' opinion that the apolipoproteins are more important than the lipids in understanding lipid transport and metabolism.

There are five classes of apolipoprotein (A–E) and many subclasses, and the apolipoproteins E have recently been sequenced. When the arginine residue is replaced at position 112 in apolipoprotein E_3, the molecule is not recognized by the lipoprotein E receptor. Homozygous inheritance of this E_3 deficiency can lead to type 3 hyperlipidaemia.

Lipoprotein receptors on cell membranes regulate entry of lipids into cells, are a target for specific lipoproteins and regulate the plasma lipoprotein concentrations. The best studied lipoprotein receptor is the low-density lipoprotein (LDL) receptor whose absence is responsible for familial hypercholesterolaemia. T.P. Bersot (University of California, San Francisco) showed that this receptor interacts with apolipoproteins B and E (Apo-B is the main apolipoprotein found in LDL whereas E is found in very low-density and high-density lipoproteins). The Apo-B/E receptor has a molecular weight of about 106,000, requires the presence of divalent cations and is inactivated by pronase.

A key difficulty in clinical research is assessing the amount of atheroma in a living human artery. D.H. Blankenhorn (University of Southern California) is using angiograms of the femoral and coronary arteries and computer image processing techniques. The centre line of the vessel is established and variations of wall distance at multiple transverse sections measured. The method has proved successful in postmortem material and is now being used *in vitro* to assess response to lipid-lowering agents. He was hopeful that the method could be applied to intravenous, rather than arterial, injections of radiographic contrast material to permit more frequent measurements. Radiological methods may be replaced by ultrasound in the not so distant future.

Lipid-lowering drugs (clofibrate, cholestyramine, nicotinic acid) have been found to be almost uniformly ineffective in reducing myocardial infarction mortality. In the WHO-sponsored clofibrate trial coordinated by M. Oliver (University of Edinburgh), there had been a reduction in non-fatal myocardial infarcts (none in fatal ones) but the overall mortality had been significantly greater in the treated group. The most effective anti-thrombotic drugs so far used had been the coumarin anticoagulants which had reduced mortality by about 20 per cent compared with placebo. Anticoagulants were almost discarded in the UK in the early 1970s but use has continued in the Netherlands (A.S. Douglas, University of Aberdeen). A recent randomized trial of withdrawal of long-term anticoagulant therapy in that country showed a higher mortality in those who stopped therapy than in those who continued therapy. Trials with anti-platelet drugs (aspirin, dipyridamole, sulphinpyrazone) had shown a smaller effect. The largest trial with aspirin had shown no benefit and the interpretation of one of the sulphinpyrazone trials was hotly contested. Overall these drugs had reduced myocardial infarction mortality by about 10–15 per cent. Four recent positive trials with β-adrenergic blockade, which had shown a reduction of 25–33 per cent in mortality during the first three years after a myocardial infarct, were mentioned only briefly. The speakers would not be drawn into a discussion of why myocardial infarction mortality has declined substantially in the USA but only very slightly in the UK. When asked whether we should all start jogging, restrict saturated fat intake, stop smoking and treat mild hypertension (or await better evidence of efficacy), one of the Americans said that we probably should, and quoted Irvine Page's aphorism "I don't want to be the smartest man in the graveyard".

The last session was concerned with platelet function and the role of prostaglandins in vascular disease. Platelet plugs form very rapidly when small vessels are ruptured and platelet adhesion and aggregation is usually the first step in arterial thrombosis. At least three different mechanisms are responsible for platelet aggregation — contact with collagen, release of ADP and release of the arachidonic acid endoperoxide product thromboxane A_2 (TXA_2). TXA_2, released from platelets and damaged tissues, has received much attention recently but G. Born (King's College, London) reported experiments on bleeding time from punctures made in small mesenteric vessles which support a major role for ADP. The question is important as great efforts are being made to develop inhibitors or antagonists of TXA_2.

PGI_2 (prostacyclin) has attracted great interest because it is formed by the endothelium of blood vessels and is a powerful inhibitor of platelet aggregation. S. Moncada (Wellcome Research Laboratories) suggested that its physiological role is to prevent interactions between platelets and the blood vessel wall. PGI_2 interacts with a receptor on platelets which raises the level of cyclic AMP, makes them less aggregable and inhibits the exposure of fibrinogen receptors on their surface. Moncada was optimistic of the usefulness of PGI_2 in preventing clotting in extracorporeal circulations (cardiopulmonary bypass, charcoal column haemoperfusion, renal dialysis), in platelet consumption coagulopathies (haemolytic uraemic syndrome) and in peripheral vascular disease (gangrene of the lower extremities, Raynauds phenomenon). Others agreed about the value of PGI_2 as a heparin-sparing agent in extracorporeal circulations but contested its value in peripheral vascular disease.

C.T. Dollery (Royal Postgraduate Medical School, London) described a new method for measuring plasma 6-oxo-PGF, the hydrolysis product of PGI_2, which uses gas chromatography-negative ion chemical ionization mass spectrometry and has a sensitivity limit of 500 pg ml^{-1}. The concentrations he measured in venepuncture samples from normal humans were very low, averaging 1.2 pg ml^{-1}, and were almost certainly too low for PGI_2 to have a physiological role in the normal circulation. Large amounts of PGI_2 produced by isolated tissues and perfused organs are caused by the injury involved in their preparation. However, PGI_2 was released in substantial quantities *in vivo* as a result of trauma caused by passing vascular catheters in man or dog and by infusions of hypertonic dextrose into a human forearm vein. He proposed that PGI_2 released from damaged endothelial cells prevented the propagation of the luminal face of a platelet clot in a puncutured vessel and inhibited

C.T. Dollery is Professor at the Royal Postgraduate Medical School, Ducane Road, London W12 0HS.

*The Biological Council symposium on 'Atherosclerosis: Mechanisms and Approaches to Therapy' was held at the Imperial College of Science and Technology, London, on 5–6 April 1982.

0028-0836/82/210286-02$01.00

platelet aggregation in a vessel that had been damaged but not severed. Possible application of this theory to atheroma depended on the frequency and importance of focal damage to arterial endothelium *in vivo*.

The symposium left mixed impressions. Animal and human studies on the pathogenesis of atheroma seem to be making only slow progress. Much of the difficulty arises from the unsatisfactory nature of the animal models and the difficulty of studying a process in man which progresses over decades. In contrast, lipid biochemistry and pharmacology are making important discoveries and eventually this basic knowledge should yield dividends in new forms of therapy. The problem is when, as there appears to be nothing of great promise on the immediate horizon. New methods of measuring plaque regression will make early human studies feasible but very prolonged outcome trials will still be needed. The general gloom about the immediate prospects for development of therapeutic agents can be summed up by two quotes from the discussion "May the risk of controlling risk be greater than the risk itself?" and "You can be sure that a new treatment will do someone harm. You cannot be sure, until you do a large-scale clinical trial, that it will do anyone any good". Your correspondent was excited by the new developments in biochemistry and pharmacology of both the lipoproteins in the blood and the fatty acids in cell membranes and hopes these pessimistic views are wrong. □

The nuclear pore — a biological grommet?

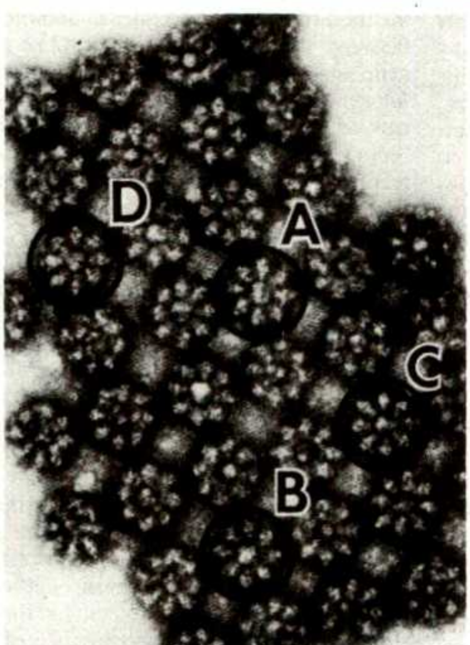
from N. Michael Green

THE double membrane surrounding the nucleus is penetrated by nuclear pore complexes, recognized from the early days of electron microscopy as large structures (1,200 Å in diameter) with a striking eightfold rotational symmetry (reviewed by J.R. Harris in *Electron Microscopy of Proteins* Vol. 1, Academic, New York, 1981). Although they are referred to as pores, the channel is usually filled with a dense plug. The general features of the complex have long been recognized but higher resolution has been prevented by the lability of the structure, which is readily distorted by the processing required for microscopy. New information has now emerged in a paper by Unwin and Milligan (*J. Cell Biol.* **93**, 63; 1982). They used the favoured nuclear envelope of the *Xenopus* oocyte, which contains such a high density of pores that they tend to form tetragonal arrays. The topology of such an array is similar to that of the intercellular gap junction, but whereas the latter shows a regular lattice stabilized by protein interactions, the pores in the nuclear envelope remain distinct and appear to have rotational freedom in the paired lipid bilayers. Nevertheless, the close packing provides mutual support and minimizes distortion caused by drying in the negative stain (either uranyl acetate or gold thioglucose).

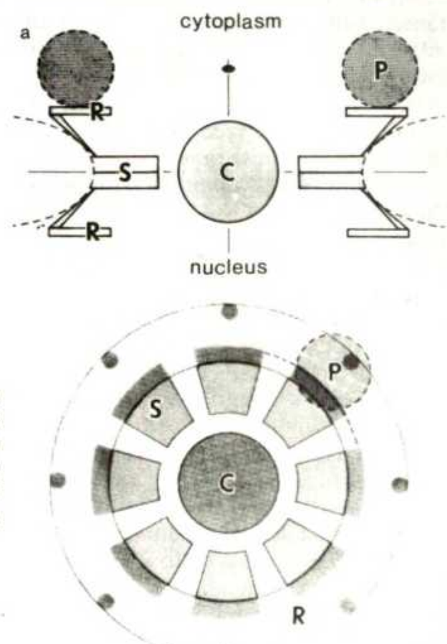
N. Michael Green is in the National Institute for Medical Research, Mill Hill, London NW7 1AA.

The most revealing pictures were obtained by allowing the nuclear envelope to adhere to hydrophilic, polylysine-treated carbon film and then washing either with high salt or with 0.1 per cent Triton X-100. This released both the intact complexes and various substructures, which adhered to the film in the neighbourhood of the original envelope. The following constituents were observed: (1) low-contrast rings with an internal diameter of 800 Å and an external diameter of 1,200 Å; (2) a dense central plug of diameter 370 Å; (3) eight particles of diameter 220 Å, possibly ribosomes, bound to the rings, enhancing their eightfold symmetry, and (4) broad radial spokes which in some way link the plug to the rings. Deciphering the three-dimensional organization of these components was greatly aided by occasional oblique and edgewise views of the complex, from which maximum information was obtained by examining the specimen at various angles on the tilting stage. Individual images, edgewise or *en face*, were analysed by numerical Fourier methods to give an accurate radial density distribution and an approximate distribution perpendicular to the membrane.

Because of the restriction to single images, rather than a regular array, and the problems of specimen preservation, the resolution was limited to 90 Å. Nevertheless, a fascinating picture emerges. The pore structure is basically a three-layered sandwich of a disk between two thin rings. The disk, consisting of a central plug and eight broad spokes, is suspended between the rings by slender extensions of the spokes. Each ring is coplanar with one half of the double membrane. Within the outer circumference of each ring the membranes approach each other, following the



Left: Detached pore complexes released onto the microscope grid from a nuclear envelope immersed in low salt medium containing Triton X-100. A,B,C and D refer to images from which projection maps were obtained.



Right: Diagram of the nuclear pore complex: (A) in central cross-section and (B) in projection down the octad axis. The major constituents are the central plug (C) the spokes (S) and the rings (R). The nuclear membrane and the ribosome (P) are shown as dashed lines.

Reproduced from *J. Cell. Biol.* **93**, 66 & 75 (1982).

0028-0836/82/210287-02$01.00

direction of the connecting strands until they fuse in the mid-line around the periphery of the spokes, at a diameter of 1,000 Å. This interpretation of the structure implies that there are no transmembrane proteins in the usual sense; the proteins of the pore complex merely provide a grommet which stabilizes the hole formed by local fusion of the paired membranes. The eight 'ribosomal' particles were attached only to the cytoplasmic face of the grommet. These particles were readily removable by high salt or EDTA, leaving a structure which at this resolution had eight twofold symmetry axes at right angles to the main eightfold axis.

These findings provoke many further questions about both the structure and function of the pore complex. It has long been known that the nuclear membrane is freely permeable to ions and small metabolites and that even molecules as large as about 90 Å in diameter can pass across with little hindrance. This is consistent with the structure proposed by Unwin and Milligan, but it is less easy to see how ribosomal particles (200 Å in diameter), assembled in the nucleus, manage to pass through the pore. It is possible that this process may be energy dependent and that some movement of plug or spokes is involved. Further structural questions concern the proteins which form the pore structure. Earlier work by Krohne, Franke and Scheer (*Expl Cell Res.* **116**, 85; 1978) has shown only two major components (MW 73,000 to 150,000) in purified pore complexes, together with several minor components. It should be possible to determine the location of at least the major constituents by immunological methods. Beyond this there is the problem of how the assembly of such large structures is controlled. From the dimensions it can be calculated that each of the major constituents (plugs, rings and spoke assembly) must contain several hundred molecules of their constituent protein subunits, giving them 'molecular weights' of 20–40 million. There is no way in which so many identical building blocks will assemble spontaneously into a precise limited structure without some guiding activity by the minor protein constituents or other cellular components.

These provocative developments provide an interesting example of the information which can be extracted from electron micrographs by modern image processing methods, even in the absence of extensive repeating structures. □

A novel gene splice in phage Mu

from Neville Symonds

GENE expression is usually thought of as being controlled through the action of proteins, such as repressors and effectors. Recently, a different mode of control has been reported whereby inactive genes are relocated by some type of recombination process to new sites where transcription takes place. Examples of recombinational control are mating type in yeast (Leupold *Nature* **283**, 811; 1980), phase variation in *Salmonella* (Zieg & Simon *Proc. natn. Acad. Sci. U.S.A.* **77**, 4196; 1980) and host range in phage Mu (Van de Putte *et al. Nature* **286**, 218; 1980). The last example refers to the unexpected finding that two types of Mu phage can be detected which differ with regard to the bacteria they can infect. The difference in host specificity is correlated with the two possible orientations observed in the G region of Mu, an invertible segment of about 3,000 base pairs. With G in one orientation, designated (+), Mu phage can adsorb to one set of hosts, which includes *Escherichia coli* K12; in the other orientation, (–), the phage adsorb to a different set, including certain *Citrobacter* and *Shigella* strains. How does the oscillation in G orientation bring about alterations in host specificity?

In general terms the answer to this question has been known for some time. With G in the (+) orientation, two Mu genes, S and U, which code for tail fibres and enable the phage to adsorb to certain hosts, are expressed; in the (–) orientation two alternative genes, S′ and U′, are expressed, so the host range is altered. The actual genetic mechanism by which this switch is accomplished has, however, only now been elucidated and is described by Giphart-Gassler and co-workers in this issue of *Nature* (p.339).

The intriguing point to the story is that, in a manner somewhat reminiscent of that used to create diversity among immunoglobulins, the variation between the S and S′ genes is brought about by joining a constant DNA sequence, S_c, adjacent to G with either of two different sequences S_v or S_v' located within G. In the (+) orientation, transcription starts at a promoter site outside G, proceeds through S_c and S_v, and then through the adjacent U gene which lies wholly within G. Tail fibres corresponding to S and U are therefore produced. In the (–) orientation the situation is reversed, the reading being in the order $S_c + S_v'$, U′, so that S′ and U′ tail fibres are made but not those corresponding to S and U.

The recombinational control which underlies the three examples mentioned above and is exemplified by the variation in host range of Mu is fundamentally different from control of protein synthesis by positive and negative feedback circuits. Such circuits are able to modulate the activity of genes and, in the extreme case, turn them off completely. Recombinational control, on the other hand, introduces a two-way switch into these circuits, allowing alternative pathways to be introduced into the system. This added flexibility is useful when only one of two possible phenotypes can be tolerated at a particular time. It is well known from experiments involving mixed infection of cells with wild-type phage and host-range phage mutants that progeny particles which receive a mixed complement of tail fibres acquired from different parents have the worst of all worlds, not being able to adsorb properly to any of the usual hosts. If the host range of a phage, such as Mu, is to be extended by providing two distinct sets of tail fibres, then it is clearly advantageous that only one set or the other should be synthesized at any stage — a two-way switch is therefore appropriate.

The precise recombinational mechanism by which the switch between mating types is instituted in yeast, where some type of gene conversion seems to occur, is different from the inversion systems used to bring about phase variation in *Salmonella* or the alternation in host range shown by Mu (and also by phage P where a closely related system operates). All these inversion switches use site-specific recombination systems which act on the short inverted repeats at either end of the relevant inverted region, and which are catalysed by genes immediately adjacent to it. Surprisingly, although the invertible regions in *Salmonella* and Mu are quite distinct, there is some sequence homology between their inverted repeats. The *gin* gene which mediates G inversion and the *hin* gene mediating inversion in *Salmonella* can complement each other (Szekely & Simon *J. Bact.* **148**, 829; 1981). The known recombinational switches of the inversion type therefore all seem to stem from some common origin. Another evolutionary oddity of the G system is that the emergence of two kinds of Mu phage with different host ranges would be expected to lead to an increase in the number of bacterial strains which harbour Mu prophage. However only the original Mu isolate detected in *E. coli* K12 has ever been reported. Whatever the reason for this, the whole of the G region has been preserved in the Mu phage we have, even though within K12 strains part of the region (and the *gin* gene) could easily be deleted and the phage locked in the (+) orientation. Perhaps there is still another twist to come in the G story by which G and *gin* are implicated in some other aspect of the Mu life cycle. □

Neville Symonds is Professor of Microbial Genetics in the School of Biological Sciences, University of Sussex, Brighton BN1 9QG.

0028-0836/82/210288-01$01.00

REVIEW ARTICLE

Active chromatin

Stuart Weisbrod

Cold Spring Harbor Laboratory, PO Box 100, Cold Spring Harbor, New York 11724, USA

Active genes are packaged into an altered nucleosome structure forming a chromosomal domain defined by increased sensitivity to nucleases. This structure, reflecting a potential for transcription, contains sites hypersensitive to nuclease digestion adjacent to the coding regions and may also be distinguished by specific non-histone proteins, variant or modified histones or modified DNA. Its formation, by unfolding of a tightly packed chromatin fibre by factors which might affect DNA supercoiling, may be the first step in gene activation.

REGULATION of gene expression is a fundamental mechanism in eukaryotic development and survival. The structure of the chromosome is directly related to this regulation since, for the most part, the DNA in the nucleus is compacted too tightly to be available to the transcriptional apparatus. Therefore the chromatin structure of genes which are transcribed, approximately 10–20% of the total, has to be different from that of the bulk of the DNA. Both the 'housekeeping' genes, those which are expressed in all cell types in the organism, and the genes which distinguish one cell type from another, exist in the active conformation.

Studies on eukaryotic chromosomes have focused mainly on the basic unit of chromatin, the nucleosome, and on how the individual nucleosomes are packed together. The existence, structure, and packaging of nucleosomes have been the subject of a number of reviews[1–3]. Most of the evidence in these studies stems from electron microscopy, where a beaded chromatin fibre is observed, and enzymatic studies which reveal the existence of a repeating unit. The nucleosome consists of 146 to 240 base pairs of DNA wrapped twice around a histone core, made up of two each of the four inner histones H2A, H2B, H3 and H4. A fifth histone, H1, is also associated with the nucleosome but in a manner which is not clear. Extensive endonuclease digestion of chromatin produces a nucleosome core particle containing 146 base pairs of DNA and the histone core (but not H1), indicating that the heterogeneity of nucleosome DNA length is due to variability in the length of DNA linking one nucleosome to the next. The digestion studies also suggest that histone H1 is associated with this 'linker' DNA.

The nucleosome is also involved in the higher orders of structure that fold the DNA into the compact form found in the nucleus. The basic level of organization is the 100 Å 'beads-on-a-string' fibre generated by folding of the internucleosomal linker DNA[4]. This fibre is then coiled into a 300 Å fibre, possibly induced by histone H1 as a crosslinker, to yield a structure with a packing ratio of 25:1, close to that of interphase chromatin[5]. To make a mitotic chromosome from the 300 Å fibre a further two orders of magnitude of compaction are generated by specific nonhistone proteins which form the skeleton or scaffold of the chromosome[6]. Electron microscopy of histone-depleted metaphase chromosomes has demonstrated that the DNA is organized in loops at least 30–40 kilobases in length. Each loop of the chromatin is attached to the scaffold at its base, either by DNA–protein or DNA–RNA interactions[7], and may exist in an expanded or a highly compacted conformation.

A nucleosomal repeat structure does not distinguish between inactive and active genes. Thus the DNA sequence complexity of mononucleosomes is virtually indistinguishable from that of total DNA and they contain sequences corresponding to transcripionally active genes[8]. The nucleosome structure of active genes has also been demonstrated by immuno-electron microscopy[9]. However, there are many ways in which the structure of chromatin from actively transcribed genes could or does differ from that of non-transcribed regions; what is known of these differences will be summarized here.

Nuclease digestion studies

Evidence that chromatin from actively transcribed genes exhibits a structure different from that of non-transcribed regions stems from the use of relatively nonspecific endonucleases as probes in anticipation that they would recognize something special about active genes. Weintraub and Groudine[10] have shown that the globin gene in chick erythrocyte nuclei is preferentially sensitive to digestion by pancreatic DNase I, but not to digestion by micrococcal nuclease. The resistance of the globin gene to micrococcal nuclease suggests that the globin gene is packaged into nucleosome-like particles; its sensitivity to DNase I indicates that these particles are conformationally different from most nucleosomes. The sensitivity of the globin gene to DNase I is tissue specific. Globin chromatin is preferentially digested in red blood cells, but not in the oviduct; conversely, the ovalbumin gene is preferentially sensitive in the chick oviduct but not in red blood cells[11].

The sensitivity to DNase I of actively transcribed genes seems to be a general phenomenon[12–15], but only reflects a potential for a gene to be transcribed rather than transcription itself. For example, the ovalbumin gene remains sensitive in the hormone-withdrawn chick oviduct[16], *Physarum* ribosomal DNA remains sensitive during mitosis when ribosomal RNA transcription is not detectable[12] and the mouse globin gene exhibits the same sensitivity in induced (transcribed) as well as uninduced (non-transcribed) Friend erythroleukaemia cells[17]. The boundaries of the sensitive regions of chromatin are very precise, since in a hamster cell line transformed by adenovirus, only those adenovirus sequences that are transcribed are sensitive to DNase I while adjacent non-transcribed sequences are not sensitive[13].

Using DNase II, an enzyme which cuts primarily between nucleosomes, Gottesfeld and Butler[18] have shown that after chromatin is digested, nontranscribed nucleosomes can be selectively precipitated with 2 mM Mg^{2+}. The DNA in the active, non-precipitated nucleosomes is enriched for sequences coding for messenger RNA. Surprisingly, the active nucleosomes sedimented at 14S compared with 11S for bulk nucleosomes. When treated with RNase, the 14S structure converts to 11S, suggesting that the difference in S values can be attributed to the presence of nascent RNA chains. Other evidence[19] confirms that this technique, as well as similar fractionation techniques using micrococcal nuclease[20], will only separate actively transcribing genes and genes which are

0028–0836/82/210289—07$01.00

transcribed at different rates might be represented in different proportions in the active fraction. Since these procedures do not result in the complete digestion of active sequences, they are potentially useful for investigating the properties of active nucleosomes, but unfortunately they have proven difficult to reproduce.

Endonucleases have also been used to investigate higher order organization of active chromatin. A very light digestion of *Drosophila* nuclei by DNase I or micrococcal nuclease revealed preferential cleavage sites in a number of heat shock loci[20]. These hypersensitive sites were originally suggested to act as boundaries of a supranucleosomal structure, but subsequent analyses (see below) have placed them within the active domain. The overall sensitivity of specific genes and adjacent noncoding regions has been measured by monitoring the disappearance of specific restriction bands after a mild DNase I digestion of nuclei[22-24]. In one case[23], a 25 to 50 kilobase region adjacent to the highly DNase I-sensitive globin coding region in erythrocytes was characterized, distinguished by a level of DNase I sensitivity intermediate between that of an expressed and a totally inactive gene. It was postulated that this domain might be equivalent to a lampbrush loop, and the DNase I sensitivity an indication that it is in a relaxed rather than a condensed configuration.

Proteins associated with active chromatin

Once the nuclease digestion experiments had demonstrated that the conformation of active nucleosomes is different from that of bulk nucleosomes, it was logical to seek the structural basis of this difference. From monitoring the proteins released by DNase I digestion and the characteristics of the various enriched fractions the suggestion arose that the active structure is due to altered or modified histones and/or the presence of nonhistone chromatin proteins in the nucleosome.

The nonhistone protein fraction most often detected associated with active nucleosomes is that of the high mobility group (HMG). The characteristics of this group have recently been extensively reviewed[25]. HMGs are highly conserved proteins with low molecular weight (<30,000), have an unusual amino acid composition containing approximately 25% basic and 30% acidic residues, and are easily eluted from chromatin in 0.35 M NaCl. They are present in amounts only approximately 5% of the nuclear DNA content and in, at most, 10–20% of the nucleosomes. They have been divided into two groups based on their solubility properties in trichloroacetic acid (TCA). HMGs 1 and 2 are soluble in 2% TCA but insoluble in 10% TCA; HMGs 14 and 17 are soluble in 10% TCA but insoluble in 20% TCA. The presence of HMGs in isolated mononucleosomes has been demonstrated by several biochemical means[26-28]. Furthermore, by the use of fluorescent antibodies, HMG-like proteins have been shown specifically to be associated with the transcriptionally active puffs in salivary gland polytene chromosomes[29].

Mononucleosomes enriched in nonhistone proteins have been prepared by limited digestion of nuclei with micrococcal nuclease[30-32]. Levy-Wilson and her co-workers[32], for example, have shown that a fraction enriched in active sequences, prepared by mild micrococcal nuclease digestion of trout testes nuclei and 0.1 M NaCl solubilization, is enriched in HMG T (analogous to HMGs 1 and 2), H6 (analogous to HMGs 14 and 17) and ubiquitin, and is depleted in histone H1. Analyses of these HMG-containing nucleosomes suggested that H6 was preferentially bound to the histone core while HMG T was associated with the internucleosomal linker DNA.

HMGs 14 and 17 can be eluted from chick erythrocyte chromatin with 0.35 M NaCl without any detectable change in the gross structure of individual nucleosomes; however, in this depleted chromatin the globin gene is no longer preferentially sensitive to digestion by DNase I[33]. Reconstitution of the depleted chromatin with either the entire 0.35 M NaCl fraction or purified HMGs 14 and 17 at a ratio of 1 mole HMG per 10 moles of nucleosomes results in the successful reconstitution of DNase I sensitivity on the globin gene (Fig. 1)[33]. Further reconstitution studies[34] have led to the following conclusions: (1) There is no tissue specificity associated with the HMGs; thus, HMGs 14 and 17 from brain nuclei are capable of restoring DNase I sensitivity to the globin gene when reconstituted with depleted red blood cell chromatin. (2) Most actively transcribed

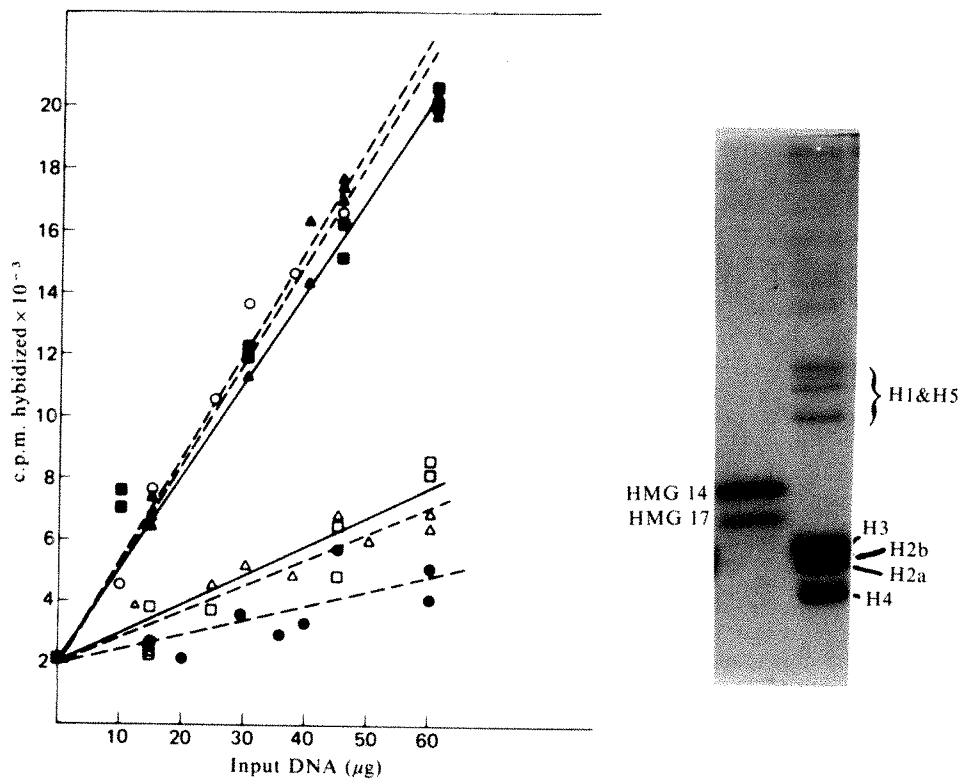


Fig. 1 Reconstitution of DNase I sensitivity of the globin gene in 14-day-old chick embryo red blood cells with 10% TCA-soluble HMG proteins. Chromatin was eluted with 0.35 M NaCl and separated into several portions. One portion was reconstituted with the 0.35 M NaCl eluate directly ▽); a second portion was reconstituted with 10% TCA-soluble fraction from the 0.35 M NaCl eluate containing mainly HMGs 14 and 17 (□); a third portion was reconstituted with the 0.35 M NaCl eluate previously treated with trypsin and then trypsin inhibitor (○); and a fourth portion was digested directly (▼). Each portion was then treated with DNase I so that 10–15% of the DNA was rendered acid soluble. The DNA was then purified and hybridized to an excess of globin cDNA. As controls, nuclei treated to 35% acid solubility with micrococcal nuclease (■) showed no preferential digestion, whereas nuclei treated to 10% acid solubility with DNase I (●) showed a significant loss of globin sequences. Reconstitution with 0.35 M NaCl alone or with neutralized and dialysed 10% TCA alone was unsuccessful in restoring DNase sensitivity (not shown). Input globin cDNA was 100,000 c.p.m. Also shown in an 18% SDS-polyacrylamide gel of total nuclear proteins and purified HMG 14 and 17 (from ref. 33).

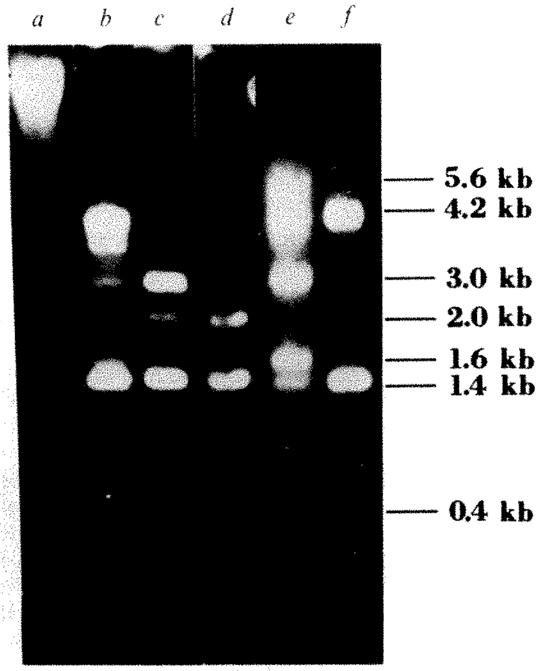


Fig. 2 The chicken adult β-globin gene is more methylated in 5-day embryos making embryonic haemoglobin than in 14-day embryos making adult haemoglobin. *a–d*, Partial digestion of adult erythrocyte DNA with *MspI*; *a*, no enzyme control; *b*, 2 units; *c*, 12 units; *d*, 100 units of enzyme; *e*, complete *HpaII* digest of 5-day erythrocyte DNA; *f*, complete digest of 14-day erythrocyte DNA. The DNA was electrophoresed on a 1% agarose gel. Southern blotted[80] and hybridized to an adult-specific β-globin probe, pHb1001. (From ref. 68.)

genes become sensitized to DNase I by HMGs 14 and 17. (3) Specificity for HMG binding seems to be in the recipient NaCl-depleted chromatin; thus, HMGs 14 and 17 from red blood cell chromatin fail to induce DNase I sensitivity to the globin gene when reconstituted with HMG-depleted chromatin from brain. (4) Either HMG 14 or HMG 17 can sensitize most actively transcribed genes to DNase I. (5) Genes transcribed at different rates have about the same affinity for HMGs 14 and 17. (6) HMGs 14 and 17 bind stoichiometrically to actively transcribed nucleosomes. (7) HMGs 14 and 17 can restore DNase I sensitivity to HMG-depleted nucleosome core particles containing 145 base pairs of DNA. And (8) as judged by sensitivity to DNase I, actively transcribed nucleosomes become associated with HMGs 14 and 17 within 3 minutes after these genes are replicated[35].

Taking advantage of the specific binding properties of HMGs described, an active nucleosome affinity column has been prepared, by coupling chicken HMGs 14 and 17 to agarose or glass beads, and used to map the DNA regions surrounding the chicken α-globin genes for their capacity to bind to HMGs 14 and 17[36]. The results indicated a direct correspondence between chromosomal regions which are capable of interacting with HMGs 14 and 17 and the regions which are highly sensitive to DNase I digestion. In contrast, the chromosomal regions of intermediate DNase I sensitivity which are adjacent to the highly sensitive regions did not bind HMGs 14 and 17.

The positioning of HMGs 14 and 17 in active nucleosomes has been investigated in a number of ways, both directly using nucleosomes enriched in active sequences, and indirectly by reconstitution. Electrophoretic studies[27,28,37,38] have indicated that HMGs 14 and 17 bind to a particle containing 150–160 base pairs of DNA, possibly replacing histone H1 and leading to a more open chromatin fibre. In contrast, HMG 14 (17)–H1 cross-linked products have been obtained[39] and it has been shown that total chick erythrocyte nucleosomes containing H1 or H5 are still able to bind HMGs 14 and 17[40]. This complication may be due to the ease with which H1 is displaced from mononucleosomes during preparation[41]. Mapping of DNase I cutting sites and thermal denaturation studies in nucleosome cores compared with HMG–nucleosome core complexes have suggested that the major sites of interaction of HMGs are located near the ends of the nucleosome core DNA[40]. Furthermore, the characterization of subnucleosome particles produced by extensive micrococcal nuclease digestion has suggested that the HMGs are (partly) associated with histone-free DNA[42]. Thus the general concensus is that HMGs 14 and 17 occupy two binding sites at each end of the nucleosome core and cover or interact with internucleosomal DNA.

Much less is known about other non-histone proteins associated with active chromatin. However, increasing attention is being paid to sequence-specific DNA binding proteins which might have as important a role in gene regulation in eukaryotes as they do in prokaryotes. For example, Engelke *et al.*[43] have isolated a protein of molecular weight 37,000, TFIIIA, from a *Xenopus* oocyte homogenate which binds to the intragenic control region of 5S DNA and appears to be required for transcription. Positive regulatory control has also recently been shown by the sequence-specific binding of the glucocorticoid–receptor complex to cloned fragments of murine mammary tumour virus DNA[44]. Conversely, repressor–operator control is shown by polyoma T antigen, which represses the synthesis of its own mRNA and the adenovirus DNA binding protein,

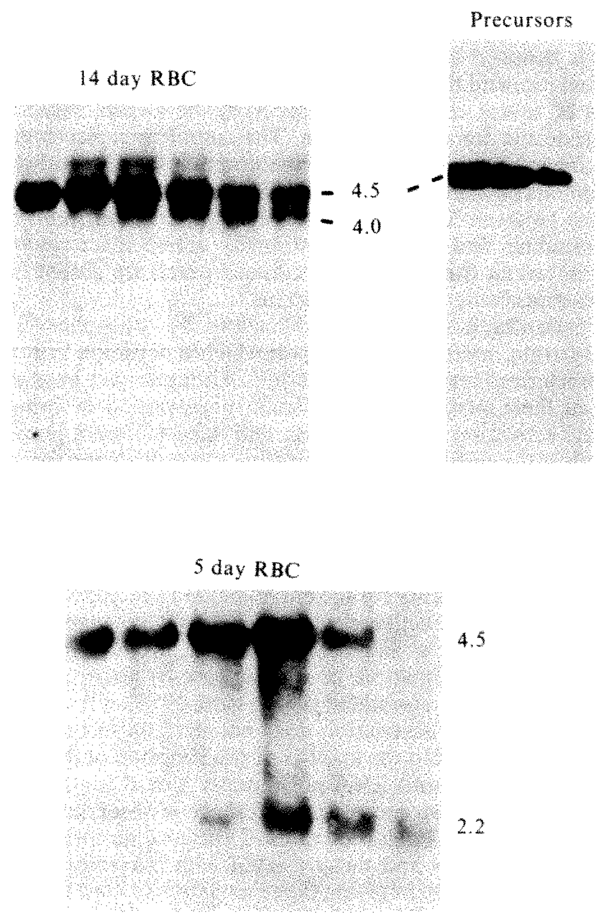


Fig. 3 Absence of DNase I-hypersensitive sites in globin chromatin from precursor cells. Nuclei from primitive erythroblasts (from 5-day embryos) making embryonic haemoglobin, definitive erythroblasts (from 14-day embryos) making adult haemoglobin, and precursor cells (500 embryos) were digested with increasing concentrations of DNase I. The DNA was purified, restricted with *Hin*dIII and blot-hybridized to an embryonic β-globin probe. Without DNase I digestion a 4.5 kilobase fragment is observed. In embryonic cells a discrete new band appears at 2 kilobases, whereas in adult cells the hypersensitive site disappears and a different region of hypersensitivity appears yielding a subfragment of about 4 kilobases. As a control the same precursor blot showed sub-bands after hybridization to a probe for the transcriptionally active actin gene. (From ref. 88.)

which represses the synthesis of adenovirus early region IV mRNA[45,46].

Modification of chromatin proteins

A fundamental manner in which chromatin conformation could be varied is by covalent modification of its histones or the DNA. Primary sequence variants of all five histones have been identified electrophoretically, especially with the use of Triton–acid–urea polyacrylamide gels[47]. Species-specific, tissue-specific and gene-specific subfractions have been described[48]. In most cases the role of histone variants in chromatin structure and function is unclear, but the occurrence of precisely timed staged-specific changes in histone subtypes during development[49,50] suggests that they are important in differentiation.

Direct evidence for the specific exchange of one histone subtype for another has been obtained by analysing male pronuclear histones from polyspermically fertilized sea urchin eggs[51]. Between fertilization and morula the sperm-specific histones of every sperm nucleus are completely replaced by cleavage stage histones. Inhibition of protein synthesis by emitine indicated that the cleavage-stage histones are stored and are of material origins. The replacement of the individual histones does not occur coordinately. The switch in histone H1 occurs almost immediately while the core histones are replaced at specific intervals during the first eight divisions. Since histone H1 is believed to play a role in the higher order structure of chromatin, and the switch occurs before any gross decondensation of sperm chromatin is observed, the switch might be a prerequisite for nuclear fusion. The complete replacement of the inner histones might serve to derepress the relatively transcriptionally inert sperm genome. Further evidence for a correlation between histone subtypes and transcription stems from the finding that the transcriptionally active macronuclei of *Tetrahymena* contains several variants which are absent in the transcriptionally inert micronucleus[52].

Postsynthetic histone modification by phosphorylation, acetylation, methylation and polyADP ribosylation is a well-characterized feature of chromatin[53]. Attempts have been made to link these modifications, particularly acetylation, to transcriptional activation and Yamamoto and Alberts[54] have proposed that histone acetylation–deacetylation can provide the flexibility needed to transduce laterally across an active gene the increased accessibility to RNA polymerase. However, even though there are numerous correlations between an increase in histone acetylation and increased RNA synthesis (ref. 55 for example), and considerable evidence for the enrichment of acetylated histones in transcriptional active chromatin[18,36,56–60], it seems unlikely that acetylation is sufficient either to produce an active chromatin conformation or to increase transcription rate. Studies of *in vitro*-assembled simian virus 40 (SV40) complexes have shown that acetylation does not destabilize or open up the nucleosome, and the kinetics of transcription of these complexes by calf thymus RNA polymerase A or B is not affected by acetylation[59]. Moreover, no difference is observed on comparing the DNase I digestion rates of nucleosome core particles from control cells and those treated with butyrate to inhibit cellular deacetylase implying that acetylation of core histones is not necessary for DNase I sensitivity[61]. So although histone hyperacetylation is a characteristic of active genes, its function, if any, in transcription, is not known.

HMG proteins also undergo postsynthetic modifications[61–65]. Recently, a careful analysis of the phosphorylation of HMGs 14 and 17 at different stages of the cell cycle in synchronized HeLa cells showed there to be a sevenfold increase in ^{32}P incorporation into HMG 14 in G_2 phase compared with G_1, and a twofold increase in incorporation into HMG 17 in early S phase[65]. In contrast, HMGs 1 and 2 were not phosphorylated. Therefore, phosphorylation of HMG 14 may be a prerequisite for the events which follow G_2—that is, chromatin condensation and cytokinesis. Perhaps of greater significance was the finding that the stage-specific phosphorylation of HMG 14 parallels that of histone H1, adding weight to the theory that HMGs 14 and 17 substitute for histone H1 in active regions.

DNA modification

The occurrence of a small fraction of modified bases in DNA has long been known (see ref. 66 for review). Over the past four years an inverse relationship has emerged between the presence of the most commonly modified base in eukaryotic DNA, 5-methylcytosine (m^5C), and gene activation. Bird and Southern[67] have shown that some type-II restriction endonucleases, whose substrate contains the dinucleotide sequence CpG, could be used to probe for m^5C by virtue of their inability to cut if the cytosine is methylated. Using these enzymes, specific sites have been shown to be undermethylated in the chicken β-globin gene in red blood cells (Fig. 2)[68], in the ovalbumin, ovotransferrin, and ovomucoid genes in the chicken oviduct[69], in amplified ribosomal genes in *Xenopus laevis*[67], in the human and rabbit globin genes[70,71], and in various integrated viral genes[72–74]. Furthermore, a direct correspondence between undermethylation, DNase I sensitivity and *in vivo* and *in vitro* transcription by endogenous RNA polymerase II has also been demonstrated[75]. A caveat to all of these correlations, though, has to be made because there are many sites within active genes which remain methylated in all tissues and other sites which remain unmethylated[69].

If undermethylation were sufficient for gene activation, inhibition of methylation should result in the expression of otherwise repressed genes. When a mouse/human hybrid cell line is grown in media containing the cytidine analogue 5-azacytidine, which is incorporated into DNA but cannot be methylated (due to the replacement of carbon 5 with nitrogen in the pyridine ring) the inactive X chromosome is reactivated[76]. Similarly, 5-azacytidine has been used to activate the metallothionein-1 gene in mouse lymphoid cells[77] and the endogenous retrovirus ev-1 in chicken lymphocytes[74]. In both cases activation has been shown to be coincident with demethylation of specific *Hpa*II sites in the coding regions of the respective gene. In the case of the retrovirus, the induction was also correlated to the acquisition of DNase I hypersensitive sites (see below) in both the 3′ and 5′ long terminal repeats. In contrast, the globin genes in the chicken lymphocytes are not

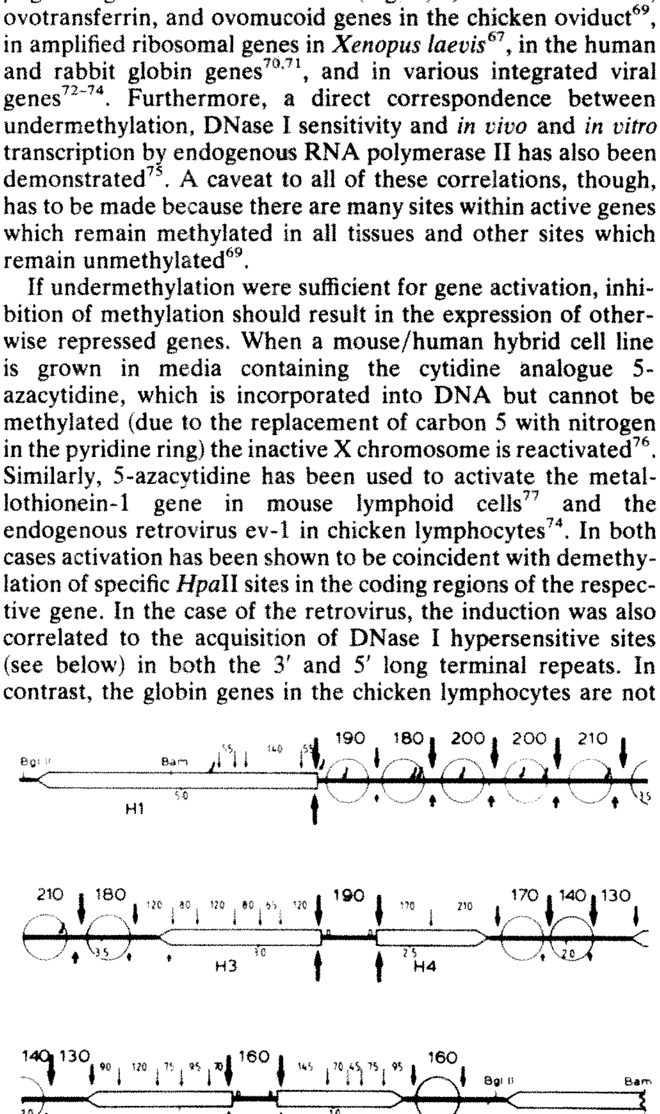


Fig. 4 Fine structure map of histone gene chromatin. Map of the 5.0 kilobase histone gene repeat, showing the transcribed regions as well as the direction of transcription for the five genes (large open rectangles with arrowheads). Numbers below the line, distance in kilobases from the *Bam*HI site on the extreme right. Small vertical rectangles above the line, TATA boxes. Arrows below the line, DNase I sites. Arrows above the line, micrococcal nuclease sites on the histone chromatin; the size of the arrows reflects the relative intensities of the corresponding bands on the autoradiograms. Curved arrows, micrococcal cleavage sites. Circles, static 146 base pair nucleosomes on the nontranscribed spacers. For clarity, 250 base pairs at the extreme left and right of the centre line are repeated on the upper and lower lines, respectively. (From ref. 84.)

activated by 5-azacytidine. Perhaps, then, the azacytidine-responsive genes are already distinct in some way from bulk chromatin. Supporting this is a finding that the inactive β-globin gene is about threefold less accessible to DNase I in *Xenopus* hepatocytes[78] than the inactive albumin gene is in erythrocytes or the vitellogenin gene is in either tissue. Therefore it seems that if demethylation is necessary for gene activation, it is not sufficient.

DNA of the transcriptionally active macronucleus of *Tetrahymena thermophila* is methylated at the N^6 position of adenine residues yielding N^6-methyladenine (m^6A)[79]. Approximately 1 in every 125 adenine residues is modified and there is no detectable m^5C. In contrast, the transcriptionally inert micronuclei contain m^5C and are devoid of m^6A. The m^6A-containing sequences are preferentially digested by both micrococcal nuclease and DNase I. Perhaps the function of these two types of modifications can be found in the differences in base pairing stability of the methylated and unmethylated pairs. The pair methyladenine·thymidine, characteristic of transcriptionally active regions, is more easily denatured than the non-methylated pair, while methylcytosine·guanine, characteristic of inactive regions, is more stable than the corresponding methylated pair.

Higher order chromatin structure

In addition to the high degree of DNase I sensitivity characteristic of active chromatin, there are also small regions of nuclease hypersensitivity, less than 200 base pairs, usually but not always 5′ to the coding region. The double-stranded cuts can be mapped by mildly digesting chromatin with either DNase I or micrococcal nuclease, purifying the DNA, redigesting with a restriction endonuclease and probing, by Southern blotting[80] with a short ^{32}P-labelled DNA fragment which abuts the restriction cut. DNA fragments smaller than the expected mapped restriction bands appear, the length of which defines the distance between the restriction site and the DNase I cut. Using this indirect end-labelling technique, Wu has demonstrated the presence of hypersensitive sites 5′ to the 70,000 and 83,000 molecular weight heat shock genes (*hsp* 70 and 83) in *Drosophila* embryos and tissue culture cells[81]. Since these sites do not exist on naked DNA, they must be a function of chromatin structure. After heat shock, the whole of the active hsp 70 coding sequence becomes sensitive to DNase I but to a magnitude less than the 5′ sensitivity, which is still partially retained. DNase I hypersensitive sites have also been demonstrated for the four small heat-shock genes[82], in the chick ovalbumin, conalbumin, globin and endogenous retroviral genes[22–24,74,75], in the rat preproinsulin gene[83], in the *Drosophila* histone gene repeat[84], and in the SV40 minichromosome[85–87]. In *Drosophila*, a pattern of micrococcal hypersensitive sites resembling that of DNase I was also observed. With both enzymes preferential cuts within the gene itself are seldom seen.

In an extensive set of experiments, the occurrence of DNase I hypersensitive sites has been correlated to β-globin gene activation in early chick development[88]. A comparison was made between a highly enriched red blood cell precursor population isolated from 20–23-hour chicken blastoderms deficient in globin transcription, embryonic red blood cells from 5-day old chicken embryos which express the embryonic β-globin gene and adult red blood cells from 14-day-old chicken embryos which express the adult β-globin gene. Hybridization of an embryonic β-globin probe to a *Hin*dIII digest of DNase-treated nuclear DNA is shown in Fig. 4. No hypersensitive sites were observed in the red blood cell precursors, whereas the embryonic and adult cells both had hypersensitive sites but different ones. It should be noted that the change in chromatin structure from embryonic to adult red blood cells does not occur directly; rather, the stem cell chooses one of two directions. In these experiments transcriptional activity was further correlated to undermethylation. Since the two globin genes exhibit the intermediate level of DNase I sensitivity in both cell types, it is attempting to conclude that the presence or absence of a hypersensitive site dictates which gene is transcribed. Other evidence suggests that this is not always true. The α and β major globin genes in murine erythroleukaemia cells show a high level of DNase I sensitivity and are transcribed at a low but detectable basal level. However, upon induction of globin synthesis with hexamethylene bisacetamide a new 5′ DBase I hypersensitive site is generated[89]. A possible problem with this result, however, is that the globin transcription detected before induction might only be occurring in a small subpopulation of cells.

Recent results have strongly suggested that the nuclease hypersensitivity of certain regions of chromatin is due to localized single-stranded regions of DNA in a nucleosome-free environment. Tissue-specific cleavages roughly corresponding to the known DNase I hypersensitive sites 5′ to the chick globin genes have been observed using the single-strand specific nuclease S_1 (ref. 90). When DNA regions containing these sites are subcloned into the pBR322 plasmid the cleavage sites are retained if the plasmid is supercoiled but lost if it is linearized. This suggests that the strain of supercoiling induces a change in DNA conformation which is sequence-specific. This supercoiling stress might be mimicking the natural effect of an altered nucleosome conformation or a sequence-specific binding protein in chromatin. Furthermore, it has recently been shown[24] that the 5′ hypersensitive site in the chick adult β-globin gene actually extends over 200 base pairs, 115 base pairs of which can be excised by *Msp*I digestion as protein-free DNA, suggesting the absence of a nucleosome.

DNA footprinting analysis of the interaction between the *lac* operator and repressor has shown that the binding domain of a protein can be 10- to 100-fold more sensitive to DNase I digestion when the protein is bound[91]. Therefore hypersensitive sites may represent regions of specific interaction with DNA sequence recognizing proteins. Consistent with this is the finding that the origin of replication and transcription of SV40 and polyoma which interacts with T antigen and the promoter region of the *Xenopus* 5S RNA genes which interact with TFIIIA[92–95] are preferentially nuclease sensitive. The presence of a specific DNA binding protein probably excludes nucleosome formation. Reconstitution studies[96] using calf thymus histones and *lac* operator DNA have indicated a preferred register of histones with respect to DNA sequence, demonstrating a further limit in the distribution of nucleosomes. Moreover, in the 250 Å chromatin fibre, the nucleosomes are arranged in a zig-zag fashion[97,98], the DNA in one nucleosome adjacent to the histones in the next enabling minor changes in position to be rapidly propagated along the fibre.

These studies have led to the postulate, and subsequent demonstration, that nucleosomes could be located in a unique or a small number of distinct positions relative to DNA sequence. This arrangement, termed phasing, has been the subject of two recent reviews[99,100]. It has led to a disagreement (perhaps semantic) of whether phasing is the result or the cause of the binding of regulatory proteins with respect to DNA sequence. There has also been some concern that phasing was an artefact of the marked specificity of micrococcal nuclease for A+T-rich regions. However, in carefully controlled experiments using both micrococcal nuclease and DNase I, it has been shown, for example, that nucleosomes are precisely positioned on the nontranscribed spacer region in the *Drosophila* histone

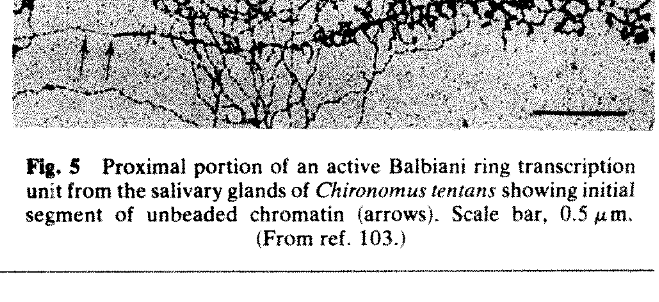


Fig. 5 Proximal portion of an active Balbiani ring transcription unit from the salivary glands of *Chironomus tentans* showing initial segment of unbeaded chromatin (arrows). Scale bar, 0.5 μm. (From ref. 103.)

gene repeat (Fig. 5)[84]. It was also shown that the nucleosome arrangement is established early in development and is retained throughout the remainder of the *Drosophila* life cycle.

Electron microscopic analysis

It is clear from the experiments described above that gene expression in eukaryotes is characterized by many levels of regulation. Most of the DNA in a nucleus, while packaged into nucleosomes, is further folded into an extremely compact and condensed higher order structure opaque to the transcriptional machinery. It might be that the features associated with active genes such as undermethylation, nuclease sensitivity, acetylation, HMGs 14 and 17 act coordinately to affect the degree in which individual genes are available to prokaryotic-type regulatory mechanisms. In the case of lampbrush and polytene chromosomes[101,102], gene activation is directly preceded by the chromatin fibre changing from a tightly packed state to a more open loop-like conformation. Specific active transcription units have been visualized for the Balbiani rings on chromosome IV in the salivary glands of *Chironomus tentans*[103] and on the ribosomal RNA genes in *Oncopeltus fusciatus* embryos[104] by an electron microscopy technique developed by Miller and Bakken[105].

Inactive chromatin exhibits a uniform beaded appearance in the electron microscope, while active transcription units are visualized as a gradient of ribonuclear protein fibres anchored to the base of the chromatin fibre by RNA polymerase molecules to yield the archetypical 'Christmas tree', the apex of which is coincident with the initiation of transcription. The Balbiani rings 1 and 2 transcription units generate 75S mRNAs coding for the salivary polypeptides. Careful measurement of the length of the transcribing chromatin has shown a marked extension of the fibre, the packing ratio going from 1.9 to 1.6 (defined as the length of B-structure DNA per unit length of chromatin)[103]. Nucleosome-size beads were observed in the active transcription units but with a frequency decreased to 4 per μm from the 28 per μm found in inactive chromatin. Furthermore, there appeared to be a smooth nonbeaded segment, 0.18 μm or about 500 base pairs in length directly preceding the RNP fibre gradient (Fig. 5). This could correspond to a 5' DNase hypersensitive region (see above), though it could also be an artefact of unusual tension during the spreading procedures.

Structural changes in ribosomal DNA transcription units, which can be identified morphologically, have been studied during early development in *Oncopeltus*[104]. At 32 hours after fertilization (late blastula) only beaded chromatin is seen, but 6 hours later, before ribosomal RNA synthesis is detected biochemically, nonbeaded stretches of ribosomal chromatin appear—many of them lacking transcripts. Again, there is a decrease in packing ratio in this case from 2.3 to 1.2. Importantly, the rDNA still exhibits a 200-base-pair repeat, though as transcriptional activity increases so does the accessibility to nucleases[106]. This result eliminates the possibility that the presence of numerous RNA polymerase molecules are responsible for the change in chromosome structure and indicates that the major change actually precedes transcriptional activation. This could occur by decondensation of a domain of the chromosome, and may possibly be equivalent to the preactivation state described by Weintraub and his colleagues[107].

Gene activation

Irrespective of the many characteristics of active genes described here, transcriptional activation still relies on the initial decondensation of a domain of the chromosome containing the gene to be transcribed. Recently, two possibly inter-related phenomena have been described which may be involved in this event. The yeast *Saccharomyces cerevisiae* can exhibit **a** or α mating type depending on the position in the chromosome where the mating type genes are located. There are three possible loci or cassettes which the genes could occupy: MAT, which is transcriptionally active, and HML and HMR, which are repressed through the action of several unlinked genes equivalently termed MAR or SIR[108]. When the cassettes are introduced into yeast on episomal plasmids by transformation they still exhibit normal regulatory control. Nasmyth and Abraham[109] have shown that the number of negative superhelical twists in a plasmid containing the HML locus is decreased when the plasmid is activated by being placed in a strain of yeast that is deficient in the MAR/SIR genes (Fig. 6). This raises the possibility that the MAR/SIR genes could function as type II topoisomerase molecules. Extrapolating back to the chromosome, decondensation could occur by the uncoiling of a 'lampbrush type' loop initiated by blocking of a sequence-or repertoire-specific topoisomerase.

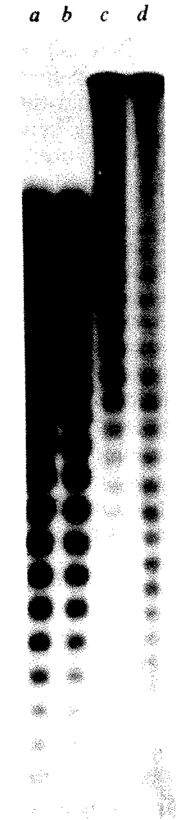


Fig. 6 Supercoiling difference in an active versus inactive yeast mating-type gene. A hybrid-yeast plasmid containing PBR, the centromeric region of chromosome 3, a selectable marker, TrP1, and either HMLα or MATa was transformed into a MAR$^+$ or a MAR$^-$ strain. The total DNA was purified and electrophoresed on a 0.7% agarose gel containing 2.5 μg ml^{-1} chloroquine. The gel was Southern blotted[9] and hybridized to pBR322. *a*, MATa in MAR$^-$ (K228 from A. Klar). *b*, MATa in MAR$^+$ (17-16 from A. Hopper). *c*, HMLα in K228. *d*, HMLα in 17-16. When the HMLα gene is repressed in 17-16 it is more negatively supercoiled than when it is active in K228. The MAT gene functions independently of MAR. (From ref. 109.)

Alternatively, decondensation/condensation may be associated with the formation of left-handed or Z-form DNA as occurs in the interband regions of *Drosophila* polytene chromosomes[110]. In physiological salt solutions Z-DNA is not stable and readily converts to the B conformation. It can be stabilized, however, by supercoiling, binding of specific basic proteins such as spermine and spermidine, or by the presence of 5-methylcytosine[111]. Rich and his colleagues[110] have proposed that selective demethylation of Z-DNA, thereby destabilizing it, would result in torsional stress. This stress can be overcome by positive supercoiling, which will result in an unwinding of the double helix at a position removed from the initial event, and yield a nuclease hypersensitive single-stranded region.

The primary events contributing to changes in higher order chromatin domains responsible for gene activation are just beginning to be investigated. Once known, secondary components directly involved in the act of transcription *per se*, such as the role of histone modification, HMGs and promoter/repressor control might be better understood. Only then may it become possible to answer some of the more fundamental questions of gene control and differentiation.

I thank Dr Ronald Laskey for suggesting that I write this review, Drs J. Hicks, K. Nasmyth, L. Silver, J. Strathern and H. Weintraub for critical comments on the manuscript, Drs B. Daneholt, J. McGhee, K. Nasmyth, H. Weintraub and A. Worcel for allowing me to reproduce figures, and M. Dermody-Weisbrod and M. Goodwin for assistance in preparing the manuscript.

1. Felsenfeld, G. *Nature* **211**, 115–121 (1978).
2. McGhee, J. D. & Felsenfeld, G. *A. Rev. Biochem.* **49**, 1115–1156 (1980).
3. Kornberg, R. D. & Klug, A. *Scient. Am.*, **244**, 52–64 (1981).
4. Thoma, F., Koller, T. & Klug, A. *J. Cell Biol.* **83**, 403–427 (1979).
5. Finch, J. T. & Klug, A. *Proc. natn. Acad. Sci. U.S.A.* **73**, 1897–1901 (1976).
6. Paulson, J. R. & Laemmli, U. K. *Cell* **12**, 817–825 (1977).
7. Benyajati, C. & Worcel, A. *Cell* **9**, 393–407 (1976).
8. Lacy, E. & Axel, R. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3978–3982 (1975).
9. McKnight, S. L., Bustin, M. & Miller, O. L. Jr *Cold Spring Harb. Symp. quant. Biol.* **42**, 741–754 (1977).
10. Weintraub, H. & Groudine, M. *Science* **193**, 848–856 (1976).
11. Garel, A. and Axel, R. *Proc. natn. Acad. Sci. U.S.A.* **73**, 3966–3970 (1976).
12. Stalder, J., Seebeck, T. & Braun, R. *Eur. J. Biochem.* **90**, 452–463 (1978).
13. Flint, S. J. & Weintraub, H. *Cell* **12**, 783–792 (1977).
14. Groudine, M., Das, S., Nieman, P. & Weintraub, H. *Cell* **14**, 865–878 (1978).
15. Garel, A., Zolan, M. & Axel, R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4867–4871 (1977).
16. Palmiter, R., Mulrihill, E., McKnight, S. & Senear, A. *Cold Spring Harb. Symp. quant. Biol.* **42**, 639–647 (1977).
17. Miller, D. M., Turner, P., Nienhuis, A. W., Axelrod, D. E. & Gopalakrishnan, T. V. *Cell* **14**, 511–524 (1979).
18. Gottesfeld, J. M. & Butler, D. J. G. *Nucleic Acids Res.* **9**, 3155–3173 (1977).
19. Gottesfeld, J. M. & Partington, G. A. *Cell* **12**, 953–962 (1977).
20. Bloom, K. S. & Anderson, J. N. *Cell* **15**, 141–150 (1978).
21. Wu, L., Bingham, P. M., Livak, K. J., Holmgren, R. & Elgin, S. C. R. *Cell* **16**, 797–806 (1979).
22. Kuo, M. T., Mandel, J. L. & Chambon, P. *Nucleic Acids Res.* **7**, 2105–2113 (1979).
23. Stalder, J., Groudine, M., Dodgson, J. B., Engel, J. D. & Weintraub, H. *Cell* **19**, 973–980 (1980).
24. McGhee, J. D., Wood, W. I., Dolan, M., Engel, J. D. & Felsenfeld, G. *Cell* **27**, 45–56 (1981).
25. Johns, E. W. *The HMG Chromosomal Proteins* (Academic, New York, in the press).
26. Albanese, I. & Weintraub, H. *Nucleic Acids Res.* **8**, 2790–2805 (1980).
27. Levinger, L., Barsoum, L. & Varashavsky, A. *J. molec. Biol.* **146**, 287–304 (1981).
28. Matthew, C. G. P., Goodwin, G. H. & Johns, E. W. *Nucleic Acids Res.* **6**, 167–179 (1979).
29. Mayfield, J. E., Serunian, L. A., Silvers, L. & Elgin, S. C. R. *Cell* **14**, 539–544 (1978).
30. Goodwin, G. H., Mathew, C. G. P., Wright, C. A., Venkov, C. D. & Johns, E. W. *Nucleic Acids Res.* **7**, 1815–1835 (1979).
31. Levy-Wilson, B. & Dixon, G. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1682–1686 (1979).
32. Levy-Wilson, B. Connor, W. & Dixon, G. *J. biol. Chem.* **254**, 609–620 (1979).
33. Weisbrod, S. & Weintraub, H. *Proc. natn. Acad. Sci U.S.A.* **76**, 630–634 (1979).
34. Weisbrod, S., Groudine, M. & Weintraub, H. *Cell* **14**, 289–301 (1980).
35. Weintraub, H. *Nucleic Acids Res.* **7**, 781–792 (1979).
36. Weisbrod, S. & Weintraub, H. *Cell* **23**, 391–400 (1981).
37. Mardian, J. K. W., Paton, A. E., Bumick, G. J. & Olins, D. E. *Science* **209**, 1534–1536 (1980).
38. Albright, S. L., Wiseman, J. M., Lange, R. A. & Garrard, W. T. *J. biol. Chem.* **255**, 3673–3684 (1980).
39. Ring, P. & Cole, R. D. *J. biol. Chem.* **254**, 11688–11695 (1979).
40. Sandeen, G., Wood, W. I. & Felsenfeld, G. *Nucleic Acids Res.* **8**, 3757–3778 (1980).
41. Renz, M., Nehls, P. & Hozier, J. *Proc. natn. Acad. Sci. U.S.A.* **74**, 1879–1883 (1977).
42. Bakeyev, V. V., Bakayeva, T. G. & Varshavsky, A. J. *Cell* **11**, 619–629 (1977).
43. Engelke, D. R., Ng, S.-Y., Shastry, B. S. & Roeder, R. G. *Cell* **19**, 717–728 (1980).
44. Payvar, F. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 6628–6632 (1981).
45. Alwine, J. C., Reed, S. I. & Stark, G. R. *J. Virol.* **24**, 22–27 (1977).
46. Nevins, J. R. & Winkler, J. J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1893–1897 (1980).
47. Alfageme, L. R., Zweidler, A., Mahowaldi, A. & Cohen, C. H. *J. biol. Chem.* **249**, 3729–3736 (1974).
48. Isenberg, I. *A. Rev. Biochem.* **48**, 159–191 (1978).
49. Cohen, L. H., Newrock, K. M. & Zweidler, A. *Science* **190**, 994–997 (1979).
50. Newrock, K. M., Cohen, L. H., Hendricks, M. B., Donelly, R. J. Weinberg, E. S. *Cell* **14**, 327–336 (1976).
51. Poccia, D., Salik, J. & Krystal, G. *Devl Biol.* **82**, 287–296 (1981).
52. Allis, C. D., Glover, C., Bowen, J. & Gorovsky, M. *Cell* **20**, 609–617 (1980).
53. Allfrey, V. G. *Chromatin and Chromosome Structure* (eds Li, H. J. & Eckhardt, R. A.) 169–191 (Academic, New York, 1977).
54. Yamamoto, K. R. & Alberts, B. M. *A. Rev. Biochem.* **45**, 721–747 (1976).
55. Pogo, B. G. T., Allfrey, V. G. & Mirsky, V. G. *Proc. natn. Acad. Sci. U.S.A.* **59**, 2239–2243 (1966).
56. Candido, E. P. M., Reeves, R. & Davie, J. R. *Cell* **14**, 105–113 (1978).
57. Sealy, L. & Chalkley, R. *Nucleic Acids Res.* **6**, 1863–1876 (1978).
58. Vidali, G., Boffa, L. L., Bradbury, E. M. & Allfrey, V. G. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2239–2243 (1978).
59. Mathis, D. J., Oudet, P., Wasylyk, B. & Chambon, P. *Nucleic. Acids Res.* **5**, 3523–3547 (1978).
60. Levy-Wilson, B., Watson, D. & Dixon, G. *Nucleic. Acids Res.* **6**, 259–273 (1979).
61. Simpson, R. T. *Cell* **13**, 691–699 (1978).
62. Caplan, A., Ord, M. G. & Stocken, L. A. *Biochem J.* **174**, 475–483 (1978).
63. Sterner, R., Boffa, L. C. & Vidali, G. *J. biol. Chem.* **253**, 3830–3836 (1978).
64. Reeves, R., Chang, D. & Chang, S. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6709–6713 (1981).
65. Bhorjee, J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6944–6948 (1981).
66. Razin, A. & Riggs, A. D. *Science* **210**, 604–610 (1980).
67. Bird, A. P. & Southern, E. M. *J. mol. Biol.* **118**, 27–47 (1978).
68. McGhee, J. D. & Ginder, G. P. *Nature* **280**, 418–420 (1979).
69. Mandel, J. & Chambon, P. *Nucleic Acids Res.* **7**, 2081–2104 (1979).
70. Van der Ploeg, L. H. T. & Flavell, R. A. *Cell* **19**, 947–958 (1980).
71. Shen, S. T. & Maniatis, T. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6634–6638 (1980).
72. Sutter, D. & Doerfler, W. *Proc. natn. Acad. Sci. U.S.A.* **77**, 253–256 (1980).
73. Desrosiers, R. D., Mulder, C. & Fleckenstein, B. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3839–3843 (1979).
74. Groudine, M., Eisenman, R. & Weintraub, H. *Nature* **292**, 311–317 (1981).
75. Weintraub, H., Larsen, A. & Groudine, M. *Cell* **24**, 333–344 (1981).
76. Mohandas, T., Sparkes, E. S. & Shapiro, L. J. *Nature* **211**, 393–396 (1981).
77. Compere, S. J. & Palmiter, R. D. *Cell* **25**, 233–240 (1981).
78. Felber, B. K. *et al. Nucleic Acids Res.* **9**, 2455–2494 (1981).
79. Pratt, K. & Hattman, S. *Molec. Cell Biol.* **1**, 600–608 (1981).
80. Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).
81. Wu, C. *Nature* **286**, 854–860 (1980).
82. Keene, M. A., Corces, V., Lowenkaupt, K. & Elgin, S. C. R. *Proc. natn. Acad. Sci. U.S.A.* **98**, 143–146 (1981).
83. Wu, C. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1577–1580 (1981).
84. Samal, B., Worcel, A., Louis, C. & Schedl, P. *Cell* **23**, 401–404 (1981).
85. Scott, W. A. & Wigmore, D. J. *Cell* **15**, 1511–1518 (1978).
86. Varshavsky, A. J., Sundin, O. N. & Bohn, M. J. *Nucleic. Acids Res.* **5**, 3469–3477 (1978).
87. Saragosti, S., Moyne, G. & Yaniv, M. *Cell* **20**, 65–73 (1980).
88. Groudine, M. & Weintraub, H. *Cell* **24**, 393–401 (1981).
89. Sheffery, M., Rifkind, R. A. & Marks, P. A. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
90. Larsen, A. & Weintraub, H. (in preparation).
91. Schmitz, A. & Galas, D. *Nucleic Acids Res.* **6**, 111–137 (1979).
92. Varshavsky, A. J., Sundin, O. & Bohn, M. *Cell* **16**, 453–466 (1979).
93. Sundin, O. & Varshavsky, A. *J. molec. Biol.* **132**, 535–546 (1979).
94. Pelham, H. and Brown, D. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4170–4174 (1980).
95. Gottesfeld, J. M. & Bloomer, L. S. *Cell* **21**, 751–860 (1980).
96. Chao, M. V., Gralla, J. & Martinson, H. G. *Biochemistry* **18**, 1068–1074 (1979).
97. Lohr, D. & VanHolde, K. E. *Proc. natn. Acad. Sci. U.S.A.* **76**, 6326–6330 (1979).
98. Worcel, A., Strogatz, S. & Riley, D. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1461–1465 (1981).
99. Kornberg, R. *Nature* **292**, 579–580 (1981).
100. Zachau, H. G. & Igo-Kemenes, T. *Cell* **24**, 597–598 (1981).
101. Gall, J. G. & Callan, H. G. *Proc. natn. Acad. Sci. U.S.A.* **48**, 562–570 (1962).
102. Beerman, W. *Chromosoma* **41**, 297–326 (1973).
103. Lamb, M. M. & Daneholt, B. *Cell* **17**, 835–848 (1979).
104. Foe, V. E. *Cold Spring Harb. Symp. quant. Biol.* **42**, 723–740 (1978).
105. Miller, O. L. Jr & Bakken, A. *Acta endocr.* **168**, 155–177 (1972).
106. Mathis, D. & Gorovsky, M. *Cold Spring Harb. Symp. quant. Biol.* **42**, 773–781 (1978).
107. Stalder, J. *et al. Cell* **20**, 451–460 (1980).
108. Herskowitz, I. & Oshima, Y. in *The Molecular Biology of the Yeast* Saccharomyces (eds Strathern, J., Jones, E. & Broach, J.) (Cold Spring Harbor Laboratories, New York, in the press).
109. Nasmyth, K. & Abraham, J. (in preparation).
110. Nordheim, A. *et al. Nature* **244**, 417–422 (1981).
111. Behe, M. & Felsenfeld, G. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1619–1623 (1981).

ARTICLES

The range and unity of planetary circulations

Gareth P. Williams & J. Leith Holloway Jr

Geophysical Fluid Dynamics Laboratory/NOAA, Princeton University, Princeton, New Jersey 08540, USA

Altering the rotation rate, obliquity and diurnal period of an Earth-like model atmosphere produces a wide range of circulation forms, some of which resemble those observed on Venus, Mars, Jupiter, Saturn and (perhaps) on Uranus and Neptune. These unified solutions suggest: that Jupiter and Saturn resemble a larger, faster-spinning Earth and possess a stress-bearing or momentum-exchanging sublayer; that easterly winds prevail in Uranus' summer hemisphere; and that Venus resembles a slowly rotating Earth if diurnal heating variations are included.

FROM Venus to Neptune, the motions of the planetary winds are being explored and subjected to theoretical analysis and prediction. On Earth, Mars and Jupiter the winds appear to exhibit similar arrangements and to obey similar laws[1–3]. These affinities stem mainly from the constraints placed on large-scale motion by planetary rotation[4,5]. To learn more about the influence of rotation on the structure of the planetary winds and to determine the range of circulation forms, we evaluate in this article the response of a representative (Earth-like) model atmosphere to changes in rotation rate, obliquity and diurnal period.

Our numerical experiments show that an Earth-like atmosphere, when placed in the rotational configuration of another planet, displays that planet's form of motion: Jupiter, Uranus

0028-0836/82/210295—05$01.00

and Venus appear to behave like rapidly rotating, oblique and slowly rotating Earths, respectively—despite thermodynamical and other differences. This suggests that the basic meteorology of the Solar System is relatively simple, unified and Earth-like, even though more complex meteorologies are needed for understanding the finer details of individual planetary circulations. The solutions thus provide an elementary indication of the dynamical laws and processes acting in the various planetary atmospheres.

Modelling

Our standard climate simulation model (GCM) integrates the equations of atmospheric motion, thermodynamics and conservation of mass and moisture in space and time, using spectral

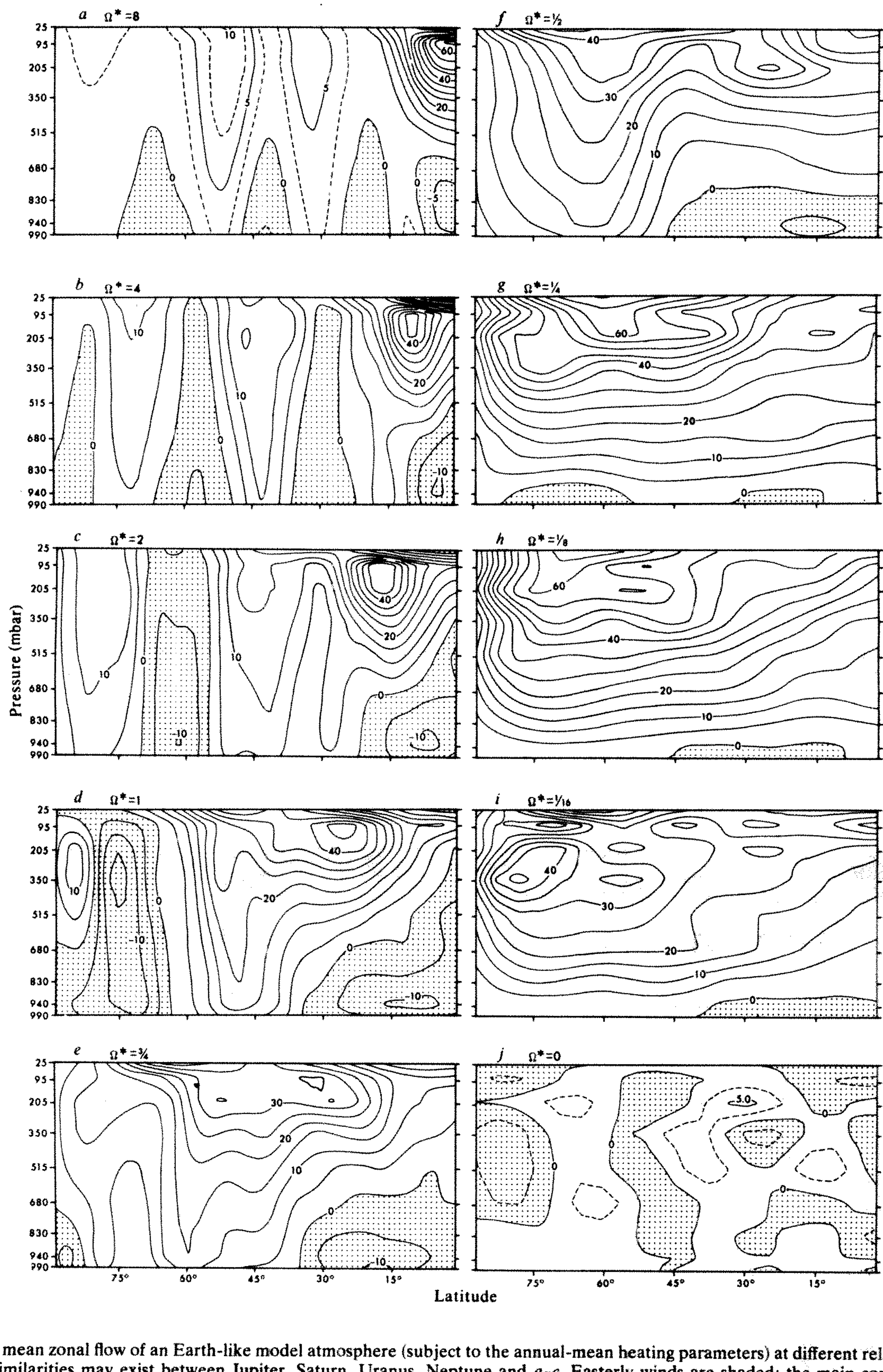


Fig. 1 The mean zonal flow of an Earth-like model atmosphere (subject to the annual-mean heating parameters) at different relative rotation rates, Ω^*. Similarities may exist between Jupiter, Saturn, Uranus, Neptune and *a–c*. Easterly winds are shaded; the main contour interval is 5 m s^{-1} and the supplementary one, 2.5 m s^{-1}. Pressure is used as the vertical coordinate. Calculations use a hemispheric sector, 120° in longitude, and horizontal rhomboidal spectral resolutions of 42 waves for the experiments in *a–e* and 15 waves for those in *f–j*.

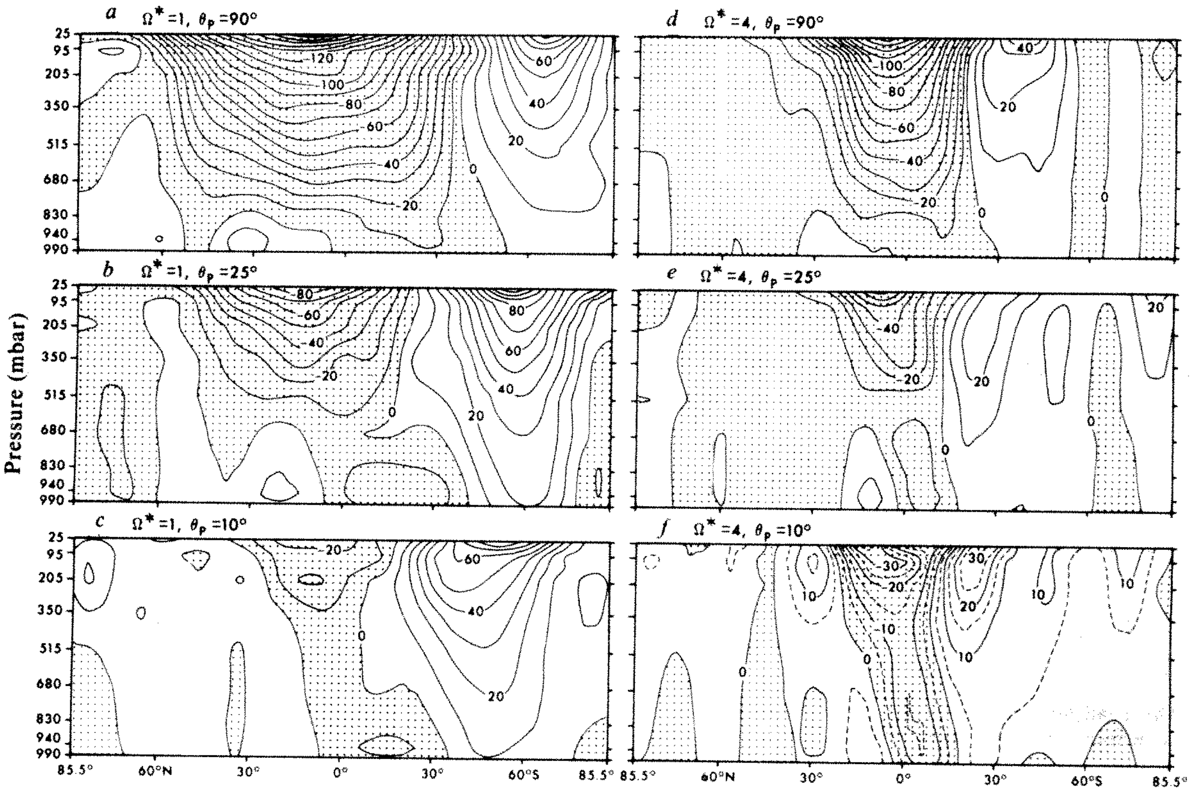


Fig. 2 The mean zonal flow of an Earth-like model atmosphere at different obliquities (θ_p) and relative rotation rates (Ω^*). Similarities may exist between Earth, Mars and *b*, between Uranus and *d*, between Saturn, Neptune and *e*. The solar constant is reduced to 3/4 or 1/2 its normal value when $\theta_p = 25°$ or 90°. The Sun is permanently at the northern hemispheric solstice (left side). Easterlies are shaded; the main contour interval is 10 m s^{-1} and the supplementary one, 5 m s^{-1}. Calculations use a global crescent, 120° in longitude, with a resolution of 15 waves (the solutions are preliminary and of reduced accuracy for the $\Omega^* = 4$ cases).

transform techniques in the horizontal and finite differencing in the vertical. For the radiative heating and cooling calculations, the normal annual-mean distribution of albedo, ozone, carbon dioxide and cloud cover are described as functions of latitude and height, while the variation of water vapour is predicted. The solar declination has its annual-mean values except when we examine obliquity and diurnal effects. For generality, the GCM has a simplified flat, uniformly moist (swamp) surface of zero heat capacity and excludes all ice-related processes. Moist convective adjustment represents small-scale convection and helps maintain vertical (hydrostatic) stability. A quadratic drag law determines the momentum and heat exchanges at the surface. (Full particulars of this GCM may be found in refs 6–9.) Solutions are discussed in terms of the relative rotation rate $\Omega^* = \Omega/\Omega_E$; where $\Omega_E = 7.292 \times 10^{-5} s^{-1}$ is the present terrestrial value. θ_p denotes planetary obliquity.

Dependence on rotation rate at equinox

Varying the rotation rate over a wide range of values ($\Omega^* = 0$–8) produces jets of diverse form and scale in the model atmospheres (Fig. 1). The jets are circumglobal and permanent in form, except at very low rotation rates ($\Omega^* < 1/64$).

The double maximum of the idealized terrestrial zonal flow stems from the overlapping of the tropical and extratropical jets at $\Omega^* = 1$ (Fig. 1*d*). These two basic jet forms are more independent and more obvious at higher rotation rates, because the baroclinic eddies are then smaller and more localized. Increased rotation rates also make the tropical jets narrower and the extratropical jets more numerous and more zonally aligned. At $\Omega^* = 8$, the tropical jet is centred on the equator (Fig. 1*a*).

In general, the mean meridional circulations consist of a direct (Hadley) cell equatorward of the tropical jet core, a weaker indirect cell in the poleward part of that jet, and assorted weaker cells in the extratropical jets. But at lower rotation rates ($\Omega^* \leq 1/4$) the baroclinic eddies cease to exist and a Hadley cell occupies the whole hemisphere (Fig. 1*g*–*j*). (The 'tropics' is defined as being that region influenced by the Hadley circulation. It extends from the equator to 15° N when $\Omega^* = 8$ and to 90° N when $\Omega^* \leq 1/4$.)

The equator-to-pole surface temperature difference generally increases with the rotation rate, from a minimum at $\Omega^* = 0$ to a maximum value (determined by the radiative equilibrium balance) approached at $\Omega^* = 8$, Table 1. A secondary minimum occurs at $\Omega^* = 3/4$ when baroclinic eddies peak in efficiency.

Dependence on obliquity at solstice

During solstice, a planet with a 10° obliquity and terrestrial rotation rate exhibits little latitudinal temperature variation, or motion, in its summer hemisphere (Fig. 2*c*). However, when the obliquity exceeds 20° the summer pole receives the greatest solar heating, resulting in strong pole-to-pole temperature gradients and extensive easterlies (with speeds up to 180 m s^{-1}) and eddies in the summer hemisphere (Fig. 2*a*, *b*). At $\theta_p = 90°$, a strong temperature inversion in the winter hemisphere excludes eddies from that region. Higher rotation rates reduce the width and amplitude of all currents and also reduce the tendency of temperature maxima to move to higher latitudes as obliquity increases (Fig. 2*d*–*f*). Consequently, the summer easterlies intrude less across the equator and the flow in the

Table 1 Surface temperatures (K) at the equator and pole of the model atmosphere as a function of relative rotation rate

Ω^*	0	1/8	3/4	1	2	4	8
T_s Equator	292	300	300	305	305	305	310
T_s Pole	280	255	275	270	265	235	230

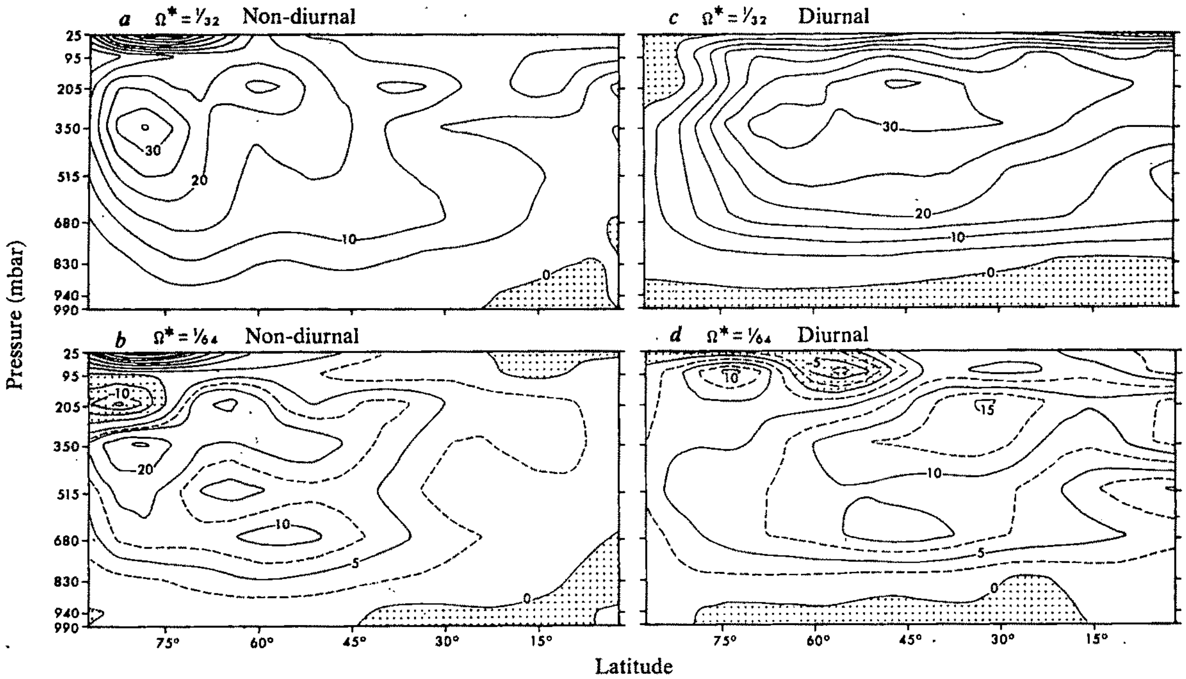


Fig. 3 The mean zonal wind of an Earth-like model atmosphere at two very low rotation rates, with and without diurnal variations in the solar heating. Similarities may exist with Venus. The main contour interval is 5 m s^{-1} and the supplementary one, 2.5 m s^{-1}. Calculations use a full hemisphere and a resolution of 15 waves.

winter hemisphere retains its equinoctal characteristics: a pre-eminent tropical jet and multiple extratropical jets. Interhemispheric heat transport, by the baroclinic eddies and the mean meridional circulation, balances radiative cooling in the unheated winter hemisphere (Table 2).

Dependence on diurnal period

The diurnal variation in the solar heating plays a fundamental role in shaping the circulation of an Earth-like planet only at low rotation rates ($\Omega^* \leq 1/16$).

As the rotation decreases towards these low rates in the nondiurnal system, the axis of the simple tropical (Hadley) jet that now constitutes the entire zonal flow moves poleward until a limit is reached at 80° latitude (Fig. 1*i*). At very low rates ($\Omega^* = 1/32$, $1/64$), geostrophy declines and the zonal flow weakens and becomes more complex (Fig. 3*a*, *b*).

Introducing diurnal variability into the very slowly rotating systems changes their narrow polar jets into broad global currents (Fig. 3*c*, *d*). The equatorial and subtropical westerly maxima (15 m s^{-1}) produced by the diurnal processes represent powerful currents for a planet whose surface rotates at <7 m s^{-1} (when $\Omega^* = 1/64$).

When $\Omega^* = 1/32$, the diurnal processes are wave-like and moderate; it may then be valid to regard the broad zonal current as being the result of the equatorward redistribution (by the waves) of momentum inherent in the polar jet of the nondiurnal state[10]. When $\Omega^* = 1/64$, the diurnal processes are so strong, and the nondiurnal jet so weak, that there may be no simple association between the diurnal and nondiurnal zonal flows[11].

Planetary implications

Earth: The hybrid mix of tropical and extratropical jets that occur at $\Omega^* = 1$ gives Earth the most complex of meteorologies. However, the thermal inertia of the oceans and ice caps reduces the strong seasonal variability normally associated with such an oblique planet. Only in the summer stratosphere are the warm polar temperatures and strong easterlies realized. The influence of obliquity may have been greater during Earth's ice-free-periods, when Ω^* and θ_p were perhaps double their present value[5], and the circulations were as depicted in Figs 1*c*, 2*b*.

Mars: Despite great differences in mass and composition, the martian and terrestrial atmospheres have similar forms of circulation. Lacking oceans, however, Mars has greater seasonal variation and, at solstice, the temperature and velocity distributions predicted by a Mars-like GCM[3] resemble those given by our simplified Earth model (Fig. 2*c*).

Venus: Venus' zonal circulation consists of a broad retrograde current that varies almost uniformly with latitude and step-like with height. Velocity maxima occur near the equator and in midlatitudes[12–14]. This form of circulation is also exhibited by our Earth-like model when the rotation is very

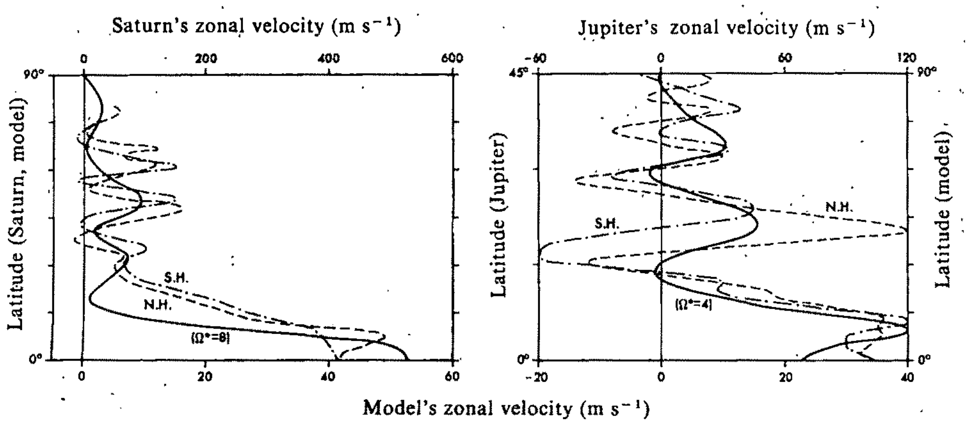


Fig. 4 Latitudinal profiles of the mean zonal velocities of northern and southern hemispheres (NH, SH) and rapidly rotating Earth-like model atmospheres. Adapted from Voyager data[17,18] and from Fig. 1*a*, *b* (at 205 mbar): *a*, Saturn and model at $\Omega^* = 8$; *b*, Jupiter and model at $\Omega^* = 4$. Scales are arbitrary.

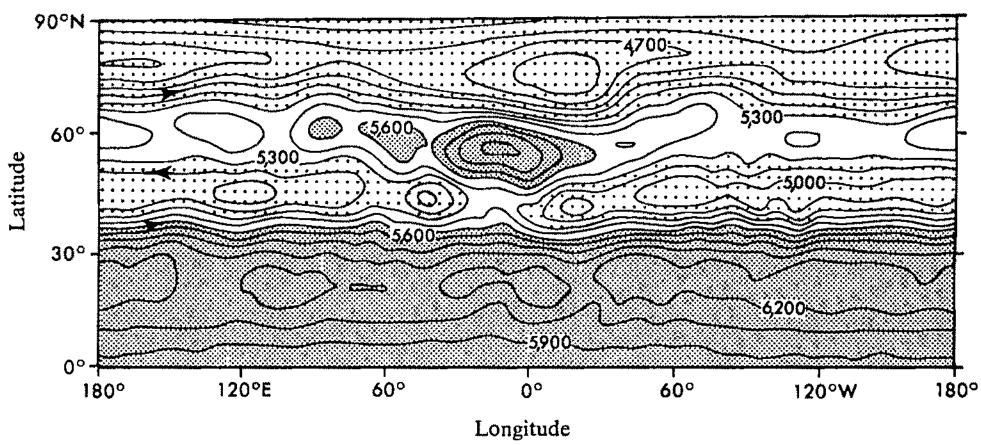


Fig. 5 The singular vortex resembling the Great Red Spot is produced by a topographic surface anomaly in a rapidly rotating ($\Omega^* = 4$) Earth-like model atmosphere. The 1-km high 'mountain' is centred near the latitude (53°) of zero mean flow (in a geostrophic, anticyclonic-shear zone) and at 0° longitude. Contours of the 500-mbar geopotential surface (a streamfunction) are plotted at intervals of 150 m with shading for values below 5,300 m and above 5,600 m.

slow and the diurnal heat cycle is included (Fig. 3*d*). Thus Venus seems to behave meteorologically like a slowly spinning, diurnally heated Earth. The diurnal heating component drives the quasi-horizontal turbulent exchanges that determine the form of the zonal wind[11] while the nondiurnal heating component drives the meridional exchanges that determine the amplitude of the zonal wind[10].

If we confine the solar heating more to the model's stratosphere, by inserting a high level opaque cloud, a more complex meridional circulation (multiple pole-to-equator cells) occurs but the zonal circulation remains basically unaltered. This suggests that the high level of heat deposition on Venus plays no fundamental role in determining the form of the zonal wind but is vital to the form of the meridional circulation. Stronger winds occur for Venus than for the model mainly because of the greater depth and density of its atmosphere[15].

Table 2 Surface temperatures (K) at summer and winter poles of the model atmosphere as a function of obliquity and relative rotation rate

Ω^*	1	1	1	4	4	4
θ_p	10°	25°	90°	10°	25°	90°
T_s Summer pole	300	310	310	280	320	310
T_s Winter pole	220	180	150	210	180	110

Reduced solar constants (Fig. 2) prevent the temperatures from exceeding 325 K at the summer pole.

Jupiter: Both Jupiter and the Earth-like GCM (when $\Omega^* = 4$) have circulations consisting of a pre-eminent tropical jet and multiple, highly zonal extratropical jets and assorted eddy fields (Fig 4*b*). If Jupiter does indeed resemble a larger, faster spinning Earth in its meteorology[2], then we predict that easterly trade winds occur in the lower tropical atmosphere and that Hadley and ageostrophic meridional circulations support the cloud band constituents.

Tropical jets cannot exist if the surface momentum exchange (drag) is excluded, so Jupiter must have a stress-bearing or momentum exchanging sublayer of some sort to be Earth-like in its tropical meteorology. The sublayer needed for tropical jet formation may not be uniform. Solitary topographic bumps (representing magnetic loops, rafts, icebergs, mountains or whatever)[16] inserted into the anticyclonic shear zones of the analogue circulations produce a long-lived anticyclonic vortex, together with a stagnant wake, a current reversal and secondary vortices (Fig. 5). This flow arrangement resembles that of the Great Red Spot (GRS). Topographic vortices of GRS scale are highly stable because they obey the planetary geostrophic (Burger) equations which, in contrast to the more familiar quasi-geostrophic equations, do not admit dispersive linear-wave solutions (Williams, G. P. and Pacanowski, R., in preparation). Thus, in this Earth-like meteorological view, the existence of the tropical jet and the genesis and permanence of the GRS all point (consistently) towards the presence of some sort of (irregular) sublayer on Jupiter. Such a possibility seems remote but is not inconceivable.

Saturn, Uranus and Neptune: Similar physical configurations should give all the outer planets similar meteorologies and, at least near equinox, zonal circulations like Jupiter's. Observations of Saturn's equatorial circulation reveal some basic differences from Jupiter that may be due to seasonal or other parametric differences (Fig. 4*a*). A velocity maximum at or near the equator, and a smaller relative amplitude to the extratropical jets, imply that Saturn resembles more an Earth-like atmosphere with $\Omega^* = 8$, than one with $\Omega^* = 4$. The greater strength and extent of Saturn's all-powerful equatorial jet may be due to the wider Hadley cell favoured by a deeper atmosphere.

At solstice, the large obliquity of Saturn and Neptune could lead to easterly winds dominating their summer hemispheres, if the internal heat sources do not reduce the seasonal variations. The extreme obliquity of Uranus and its lack of internal heat sources make the onset of a polar hot spot and powerful easterly winds and eddies inevitable in its summer hemisphere.

Conclusions

The element of unity seen in the above set of solutions suggests that relatively simple principles govern all the known planetary circulations and that these laws can be deduced best by evaluating a comprehensive atmospheric model rather than by trying to reason from individual facts of nature. The extent of planetary unity can be explored relatively simply: by looking for easterly winds deep in Jupiter's tropics or in the summer hemispheres of the other major planets.

We thank S. Manabe for guidance, and B. G. Hunt for stimulating correspondence. The study was initiated while G.P.W. was a visitor at the Department of Applied Mathematics and Theoretical Physics and Darwin College, Cambridge. P. Tunison, W. Ellis, M. Zadworny drafted the figures; B. M. Williams typed the manuscript.

Received 1 November 1981; accepted 23 March 1982.

1. Lorenz, E. N. *WMO Monogr.* **218,** 1 (1967).
2. Williams, G. P. *Nature* **257,** 778 (1975); *J. atmos. Sci.* **35,** 1399 (1978); **36,** 932, 1409 (1979).
3. Leovy, C. B. *A. Rev. Astr. Astrophys.* **17,** 387 (1979).
4. Rhines, P. B. *A. Rev. Fluid Mech.* **11,** 401 (1979).
5. Hunt, B.G. *Nature* **281,** 188 (1979); *J. atmos. Sci.* **36,** 1392 (1979).
6. Manabe, S., Hahn, D. G. & Holloway, J. L. Jr *WMO/GARP Publ. Ser.* **22,** 41 (1979).
7. Holloway, J. L. & Manabe, S. *Mon. Weath. Rev.* **99,** 335 (1971).
8. Gordon, C. T. & Stern, W. *WMO/GARP Publ. Ser.* **7,** 46 (1974).
9. Bourke, W. *et al. Meth. Computat. Phys.* **17,** 267 (1977).
10. Gierasch, P. J. *J. atmos. Sci.* **32,** 1038 (1975).
11. Rossow W. B. & Williams, G. P. *J. atmos. Sci.* **36,** 377 (1979).
12. Marov, M. Y. *A. Rev. Astr. Astrophys.* **16,** 141 (1978).
13. Schubert, G. *et al. J. geophys. Res.* **85,** 8007 (1980).
14. Rossow W. B. *et al. J. geophys. Res.* **85,** 8107 (1980).
15. Leovy, C. B. *J. atmos. Sci.* **30,** 1218 (1973).
16. Hide, R. *Nature* **190,** 895 (1961).
17. Smith, B. A. *et al. Science* **212,** 163 (1981).
18. Smith, B. A. *et al. Science* **215,** 504 (1982).

Size and shape in raised mire ecosystems: a geophysical model

H. A. P. Ingram

Department of Biological Sciences, The University, Dundee DD1 4HN, UK

Raised mires are ecosystems in which waterlogged peat accumulates above the level of the surrounding stream system. It has been suggested that waterlogging is maintained by matric forces, but a model involving impeded drainage is in better accord with the structure of the peat and with basic tenets of soil physics. At one site from which enough hydrological and soil physical data are available to conduct a preliminary test, the elliptical shape and proportions of the mire surface profile are in agreement with this model.

RAISED mires or raised bogs[1] which were distinguished from fens by Lesquereux in 1844 [2], were among the earliest types of peatland ecosystem to be recognized, and are known to be one of the world's most widespread types of mire[3,4]. Their development is marked by the accumulation of peat deposits which, in surface profile, resemble inverted clock glasses[5]. The growth of such a deposit is made possible because it remains perennially waterlogged almost to its upper surface: a feature as paradoxical as it is fundamental, seeing that the central parts of the surface may lie at altitudes up to 10 m above the marginal lagg streams into which the deposit drains. I shall consider here how saturation is maintained in the cupola-like peat deposits of raised mires and, having identified impeded drainage as the mechanism, shall then discuss implications for the shapes of these ecosystems.

Capillary hypothesis

Moore and Bellamy[6] suggested that domed peat deposits retain their water by matric forces (capillarity), while Gosselink and Turner[7] believed that 'capillary action' draws water up into raised mires. However, Granlund[8] had earlier shown experimentally that the zone of saturation would not rise beyond about 0.5 m in *Sphagnum* peat from the higher parts of a raised mire south of Stockholm and this roughly agrees with the maximum rise of 0.3–0.4 m found in a similar situation near Leningrad[9].

Application of the capillary model of soil pore structure[10]

$$h_c = \frac{2\gamma \cos \alpha}{\bar{r} \rho_w g} \qquad (1)$$

(where h_c is the height of capillary rise; γ the surface tension of water; α the contact angle, taken to be zero; $\bar{r}$ the mean pore radius, ρ_w the density of water; g the acceleration due to the Earth's gravity) suggests mean pore radii of the order of 30 μm. But to account for a deposit exceeding 5 m in height, the mean pore radius would need to be <3 μm, implying complete alteration (maximal humification[11]) of peat right up to the surface, rather than the more moderate humification generally experienced.

The capillary hypothesis is also unsatisfactory on theoretical grounds. First, it suggests no mechanism for generating a consistent shape, lower at the edges than the centre. Thus it fails to explain the clock-glass profile. Second, it misinterprets the nature of saturation in these deposits. Soil horizons where saturation is maintained by matric forces contain water under tension: they experience negative hydraulic potentials such that, if all water in a raised peat deposit were retained in this way, the water table or surface of zero potential displayed in a dip well would lie at the level of the lagg streams (Fig. 1, A). However, pools frequently occur on raised mire surfaces and, even in their absence, water tables commonly lie within a few tens of centimetres of the surface (Fig. 1, B). Most of the deposit accordingly experiences a positive hydraulic potential. Any saturated region above the water table is merely a shallow capillary fringe.

Soil structure

Pedologically, an intact raised mire where the peat-forming vegetation has not been destroyed by reclamation or peat extraction may conveniently be described as 'diplotelmic'[12]. It comprises two layers of material, namely a core of humified peat known as the 'catotelm' which occupies most of the deposit, overlain by a thin (25–50 cm) 'acrotelm' or 'active layer'[13–15]. Most of the soil's biological activity and exchange of matter and energy are concentrated in the acrotelm[16]. In Holarctic raised mires the top of the acrotelm comprises growing plant material, especially *Sphagnum*, which undergoes alteration by humification below, where it becomes transformed into the partly colloidal material of the catotelm. Because colloidal peat undergoes irreversible alteration on de-watering[17], a condition of the stability and growth of raised mires is that their water tables must stay sufficiently close to the surface to maintain perennial saturation of the entire depth of the catotelm. Thus the thickness of the acrotelm and the maximum depth of the water table (Fig. 1, C and WT) are roughly the same[18], seldom exceeding ~0.5 m.

Hydraulics and hydrology

Because matric forces alone are inadequate, it will be clear from Fig. 1 that the body of water saturating the catotelm is inherently unstable. Without atmospheric precipitation it would in time disperse either by evaporation or else sideways into the lagg streams by seepage through the peat. Water bodies of this kind, maintained 'by impeded drainage', or, more accurately, through a dynamic equilibrium between recharge and seepage, are common both in artificially drained land[19] and in more natural situations[20]. They are conveniently termed 'groundwater mounds'[21]. When developed in a homogeneous, isotropic porous medium through which water flows in accordance with Darcy's law

$$\mathbf{v} = -K \text{ grad } \phi \qquad (2)$$

(where $\mathbf{v}$ = bulk flux density; K = hydraulic conductivity; ϕ =

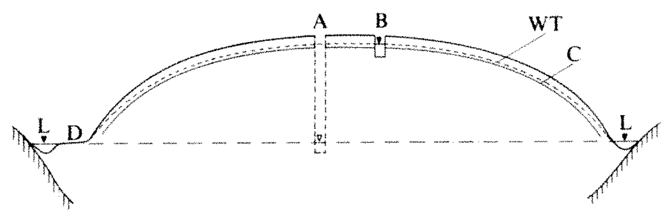


Fig. 1 Schematic sectional profile through a raised mire. A, B, dip wells (see text). C, boundary between top of catotelm and bottom of acrotelm. D, surface of seepage (a phreatic surface at which water emerges) supporting fen vegetation in the lagg. L, lagg streams. WT, water table.

0028–0836/82/210300—04$01.00

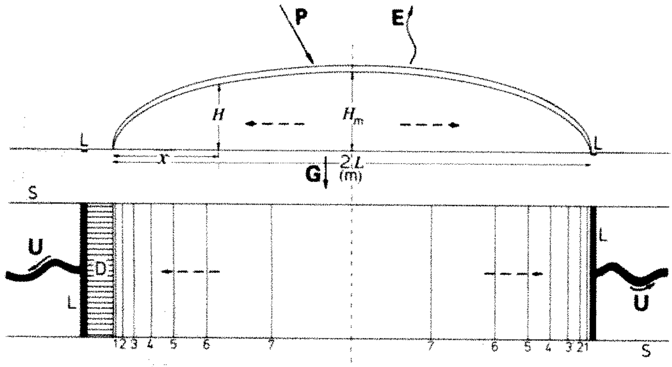


Fig. 2 Groundwater mound in a raised mire, in sectional profile (above) and in plan (below). D, L as in Fig. 1; m, summit and catchment boundary between the two lagg streams; S, parallel sides of confining valley with level floor; 1–7, contours of the water table (equipotentials according to the Dupuit–Forchheimer approximation) spaced at equal intervals of altitude. Thick arrows and bold letters: fluxes of water and flux densities (equation (5)). Thin arrows: dimensions of groundwater mound (equations (3) and (4)).

hydraulic potential), the potential distribution must conform with Laplace's equation which can be used, where the geometry is simple, to predict the relationship between height and width of the groundwater mound for any combination of discharge with hydraulic conductivity[22]. The shape of the mound can also be approximated using Dupuit–Forchheimer theory, which assumes that flow is approximately horizontal, that the equipotentials are vertical surfaces and that grad ϕ is given by the slope of the water table. Thus Childs[23] gave solutions which apply to mounds with equipotentials that are parallel, circular or elliptical in plan. The first case is simplest. Its application to an idealized deposit, shaped like a raised mire and confined in a parallel-sided valley, is shown in Fig. 2. Here the net recharge appears as **U**, the lateral discharge by seepage towards the lagg streams, which are assumed to be at bed level and of negligible depth. The permeability of the deposit is characterized by K and the quotient $\mathbf{U}/K$ is related to the dimensions of the groundwater mound by the expression

$$\frac{\mathbf{U}}{K}=\frac{H^2}{2Lx-x^2} \tag{3}$$

which takes the form of the equation of an ellipse. By setting $x=L$, the maximum height H_m of the mound is given by

$$\frac{\mathbf{U}}{K}=\frac{H_m^2}{L^2} \tag{4}$$

which corresponds to one limit of the exact solution of Laplace's equation.

Table 1 Water balance, Dun Moss, May 1972–April 1973 [26]

Method	Item	Flux density (mm yr^{-1})
Array of standard gauges, diameter 127 mm, exposed at 305 mm above ground level	**P**	530*
High water tables: lysimeter, checked against U.S.W.B. Class A evaporation pan. Low water tables: short term mass balance when **U** = 0 (dry periods)	**E**	317
Short term mass balance when **U** = 0 (dry periods)	**G**	26†
By difference	**U**	187‡
Selection of appropriate period	ΔW	0§

* Compares with $\bar{\mathbf{P}}=951$ mm yr^{-1} for decade May 1970–April 1980.
† Mott's revised estimate.
‡ Equivalent to 5.9×10^{-7} cm s^{-1}.
§ Water levels same at start and finish.

Supposing that the trough of Fig. 2 has a leaky floor, the groundwater discharge **U** becomes one item in a water budget written

$$\mathbf{P}-\mathbf{E}-\mathbf{U}-\mathbf{G}-\Delta W=0 \tag{5}$$

in which all terms are flux densities except ΔW and where **P** = precipitation, **E** = evapotranspiration, **G** = leakage and W = storage. (**G** ≪ **U** for valid application of the Dupuit–Forchheimer approximation).

Groundwater mound hypothesis

These considerations lead to the hypothesis that shape and size in raised mires are controlled by soil physics and hydrology. The hypothesis has the following elements: (1) a diplotelmic soil structure, (2) with perennial saturation of the catotelm, (3) maintained by a groundwater mound, (4) whose dimensions are governed by the water balance, (5) and by the permeability characteristics of the peat, (6) and whose surface, the water table, is confined within a thin acrotelm, (7) so that both acrotelm and ground surface assume an elliptical profile.

It follows from elements (2) and (3) that, with a thin capillary fringe, the catotelm will be coextensive with the groundwater mound when the latter is at its smallest. From elements (4) and (5) it further follows that the critical water balance is that of the driest period through which the mire survives without irreversible desiccation, when the permeability characteristics in question are those of the catotelm alone.

Testing the hypothesis at Dun Moss

To qualify as a test site for the groundwater mound hypothesis a raised mire must meet several criteria. First, human interference with its ecology and hydrology should be minimal. Second, its situation should resemble the model of Fig. 2. Third, its catotelm should have known water transmission characteristics. Finally, relevant hydrometeorological records should be available for several years, including an exceptionally dry period.

Few raised mires are likely to fulfil these stringent requirements completely, but they are met in part by Dun Moss, a small raised mire in east Perthshire (National Grid ref. NO167 558)[24]. The situation of the mire may be inferred from

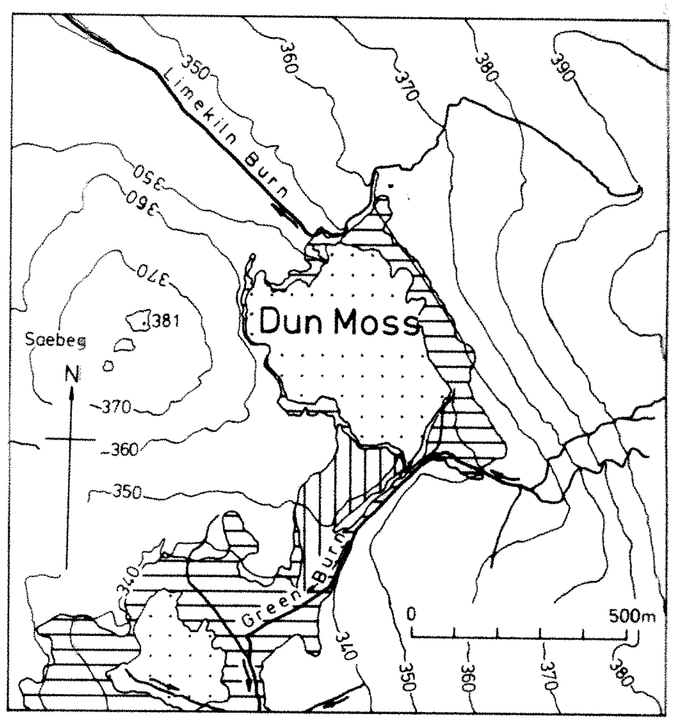


Fig. 3 Physiography and mire types in the vicinity of Dun Moss. ▩, Mire expanse; ▤, marginal lagg fens and so on; ▥, lines, area formerly exploited by peat cutting; arrows, perennial surface streams. Altitudes in m above OD, reproduced from the Ordnance Survey map, Crown copyright reserved.

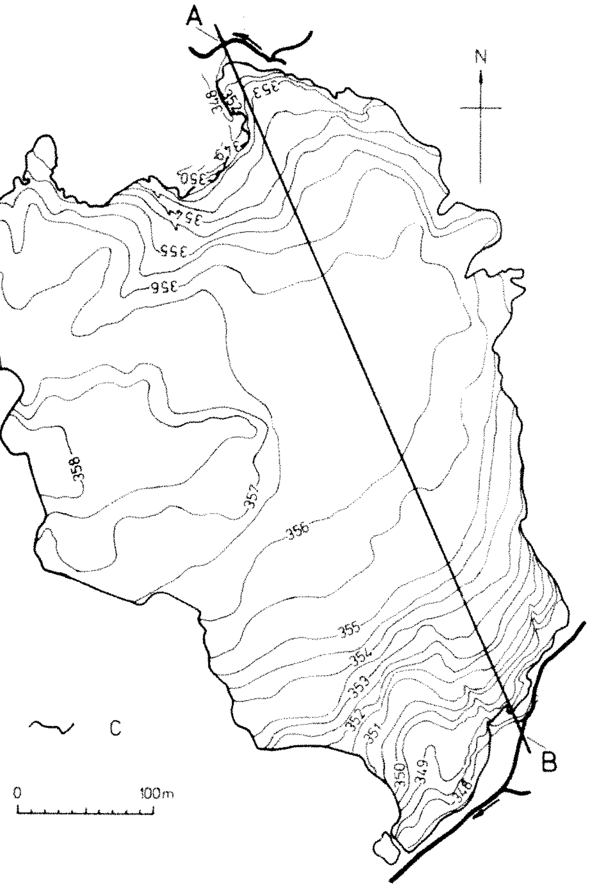


Fig. 4 Physiography of the mire expanse at Dun Moss. AB, line of profile transect connecting the captured headwaters of the Green Burn with the SE–NW reach of the Limekiln Burn. C, boundary of mire expanse. Arrows, lagg streams and so on. Contours in metres, based on a theodolite survey of the hollows on the uneven surface.

Fig. 3. It lies in a north-west trending valley with roughly parallel sides, whose headwaters were deflected south-west across the low ridge south of Saebeg through an overflow channel: a fluvioglacial feature probably eroded in late Weichselian times. Lagg streams now drain northwards to the Limekiln Burn and southwards to the Green Burn, both with beds at 348.0 m altitude where they touch the mire boundary. The surface of the raised peat deposit is a mosaic of *Sphagnum* hummocks with intervening hollows. Its general shape is shown by the contours in Fig. 4. The profile transect AB is drawn along the axis of the valley, which stratigraphical studies[25,26] have shown to continue beneath the peat deposit. On this transect the highest point of the mire surface lies at an altitude of 356.2 m, whence $H_m = 8.2$ m. The contour intersections are plotted in Fig. 5. The distance along AB between the lagg streams is 547 m, but of this 27 m is occupied by surfaces of seepage supporting *Carex rostrata* fern communities, mainly adjacent to the Green Burn. Taking 520 m as the correct value of $2L$ for substitution in equation (4) gives a morphologically based estimate of $\mathbf{U}/K = 9.95 \times 10^{-4}$ which, when substituted in equation (3), generates the thick curve drawn in Fig. 5. Except near the ends, where the Dupuit–Forchheimer approximation is likely to be weakest, and around the 100-m point on AB where the contours indicate flow diverging from the transect, the points lie close to the hypothetical ellipse.

The site has also yielded data for a preliminary estimate of $\mathbf{U}/K$ based on hydrodynamics. There are at present no sophisticated criteria for choosing the time period most appropriate for an estimate of $\mathbf{U}$, but 1 yr is the shortest period during which all the relevant seasonal biological and geophysical processes are represented and which can be selected so that $\Delta W = 0$. Since investigations began at Dun Moss, the driest year experienced has been May 1972 to April 1973, for which a provisional water balance was drawn up by Mott[26] as shown in Table 1. This suggests 5.9×10^{-7} cm s^{-1} as a tentative minimal estimate of $\mathbf{U}$. Extensive field studies of water transmission through the top of the catotelm at Dun Moss were carried out by Rycroft[25,27,28]. The results of numerous tests with seepage tubes[29] suggested that Darcy's law is only an approximate description of flow in humified peat, but that K values determined in the latter parts of long tests ($t \to 10^3$ s) give a reasonable indication of the permeability characteristics. Stratigraphical analysis showed that the raised peat deposit comprised varying amounts of the remains of *Eriophorum vaginatum* and *Calluna vulgaris* in a matrix of *Sphagnum* species (mostly *S. imbricatum* and sect. *Acutifolia*), generally humified to about H4 on von Post's scale. Consistent sets of results were obtained from H4 *Sphagnum* peat with traces of *Eriophorum* at depths of 70–80 cm at sites C510F and C520F: neighbouring positions within 60 m of the centre of the mire expanse. Variable head tests showed $\bar{K} \to 2.1 \times 10^{-4}$ cm s^{-1} while constant head depletion tests showed $\bar{K} \to 7.9 \times 10^{-4}$ cm s^{-1}, suggesting 5.0×10^{-4} cm s^{-1} as a modal value which we may tentatively accept for the catotelm as a whole, there being no seepage tube technique yet available for use at peat depths exceeding about 1.5 m. Taking these tentative values as best available estimates, we obtain

$$\mathbf{U}/K = 1.2 \times 10^{-3}$$

as a hydrodynamically based estimate of the quotient for the driest year experienced. This exceeds the morphologically based estimate by 20% and corresponds with the thin curve in Fig. 5.

Discussion

Ecosystems are entities in which biological, geochemical and geophysical processes interact. Conceptual models of ecosystems may, as with feeding or foraging studies, be fundamentally biological. When concerned with nutrient cycling they are basically geochemical. When dealing with energy capture or water movement their essence may be geophysical. All three elements may be involved in models of phenomena like succession and may contribute to our understanding of whole ecosystems[30]. Peatlands owe their existence to peculiar soil water regimes. It follows that we should look especially to soil physics and hydrology in our early efforts to build conceptual models of mire ecosystems.

In the present instance, precise conformation between the profile of Dun Moss and its hydrodynamics is scarcely to be expected. Among other complications, it is unlikely that our estimate of $\mathbf{U}$ for the driest year for which we happen to have data is actually the lowest discharge experienced. Furthermore the deposit merely resembles the model envisaged in Fig. 2, and its geometry in plan suggests that seepage is neither rectilinear nor parallel to the valley axis throughout the groundwater mound. However, the similarity of the profile to an ellipse whose parameters are in reasonable accord with what is known

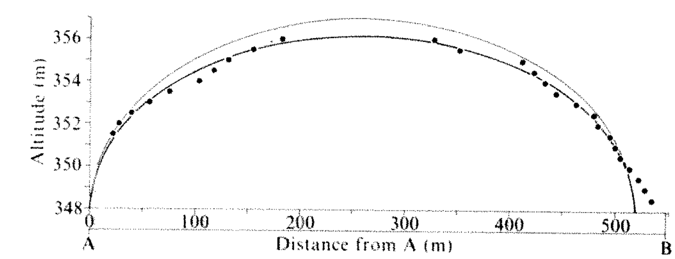


Fig. 5 Profile of mean hollow altitudes on the surface of Dun Moss in the vicinity of transect AB (Fig. 4). Plotted points: contour intersections from Fig. 4. Thick line, Dupuit–Forchheimer ellipse for $\mathbf{U}/K = 9.95 \times 10^{-4}$, the morphometric estimate, denoting the shape predicted on the groundwater mound hypothesis. Thin line, the same for $\mathbf{U}/K = 1.2 \times 10^{-3}$, the hydrodynamic estimate for the driest year in a decade. (Vertical scale exaggerated 20-fold.)

of the hydrodynamics of the catotelm seems to support the groundwater mound hypothesis.

This hypothesis accords with work by Weber[31], whose early systematic studies of raised mire ecosystems led him to conclude that they were sustained by meteoric water and impeded drainage. It also postulates a mechanism for the close control of the shapes of these ecosystems. Given that the permeability of moderately altered *Sphagnum* peat (H3–H6) varies rather little with humification[32], equation (4) predicts that damper climates will support raised mires of more pronounced convexity (higher ratios of $H_m:L$). Such trends have been reported for Southern Sweden[33] and the Federal German Republic[34], while a comparison of the convexity of raised mires in two regions of Finland with similar climates showed little difference[35]. Equation (3) suggests that growth at all points on the surface of a raised mire is likely to proceed at synchronal rates, even under fluctuating climates. This agrees with stratigraphical findings such as those of Aaby[36], and with evidence from pollen analysis of cores from various parts of Dun Moss[26] which suggests that its shape has remained remarkably constant during some 5 millennia of accumulation.

As a conceptual model, the groundwater mound hypothesis has a secure physical foundation and seems consistent with evidence from various sources. To explore it further we need data on the permeability of the deeper catotelm and on groundwater discharge as a water budget item, not only from Dun Moss but also from other raised mires of relatively simple geometry.

I thank P. J. Mott for permission to quote his water balance computation; R. G. Crawford for help in compiling Fig. 4; D. J. A. Williams for discussions; owners and tenants D. W. Rycroft, R. S. Clymo and Olivia M. Bragg for comments and for access to the site. The data from Dun Moss were obtained during work supported by the NERC.

Received 28 January; accepted 26 March 1982.

1. Ganong, W. F. H. *Bot. Gaz.* **16**, 123–126 (1891).
2. Früh, J. & Schröter, C. *Beitr. Geol. Schweiz. geotech. Ser.* **3**, 12–13 (1904).
3. Moore, P. D. & Bellamy, D. J. *Peatlands*, 10–37; 185–188 (Elek, London, 1973).
4. Kivinen, E. & Pakarinen, P. *Ann. Acad. sci. fenn.* **AIII**, 1–28 (1981).
5. Eggelsmann, R. & Schuch, M. in *Moor- und Torfkunde* (ed. Göttlich, K.) 153 (Schweizerbart'sche, Stuttgart, 1976).
6. Moore, P. D. & Bellamy, D. J. *Peatlands*, 18 (Elek, London, 1973).
7. Gosselink, J. G. & Turner, R. E. in *Freshwater Wetlands* (eds Good, R. E., Whigham, D. F. & Simpson, R. L.) 73 (Academic, New York, 1978).
8. Granlund, E. *Sver. geol. Unders. Afh.* C(373) 16–20 (1932).
9. Romanov, V. V. *Hydrophysics of Bogs* (transl. Kaner, N.) 44–59, 69 (Israel Programme, Jerusalem, 1968).
10. Hillel, D. *Soil and Water: Physical Principles and Processes*, 185–186 (Academic, New York, 1971).
11. Burghardt, W. *Geol. Jb.* **F4**, 85–86 (1977).
12. Ingram, H. A. P. *J. soil Sci.* **29**, 224–227 (1978).
13. Lopatin, V. D. *Vest. leningr. gos. Univ.* **2**, 45–46 (1949).
14. Ivanov, K. E. *Gidrologiya bolot.* 160–161 (Gidrometeoizadat, Leningrad, 1953).
15. Ivanov, K. E. *Water Movement in Mirelands* (transls Thomson, A. & Ingram, H. A. P.) 57–60 (Academic, London, 1981).
16. Romanov, V. V. *Hydrophysics of Bogs* (transl. Kaner, N.) 21–22 (Israel Programme, Jerusalem, 1968).
17. Hooghoudt, S. B. *Trans. 4th int. Congr. soil Sci.* Amsterdam 2, 31–34 (1950).
18. Ivanov, K. E. *Water Movement in Mirelands* (transls Thomson, A. & Ingram, H. A. P.) 59 (Academic, London, 1981).
19. Luthin, J. N. *Drainage Engineering*, 151 (Wiley, New York, 1966).
20. Marino, M. A. *Wat. Sci. Engng Pap., Univ. Calif.*, Davis (2004), 1 (1975).
21. Marino, M. A. *J. Hydrol. Amsterdam* **22**, 295–301 (1974).
22. Polubarinova-Kochina, P. Ya. *Theory of Groundwater Movement* (transl. de Wiest, J. M. R.) 281–282 (Princeton University Press, 1962).
23. Childs, E. C. *An Introduction to the Physical Principles of Soil Water Phenomena*, 338–340, 406–408 (Wiley, London, 1969).
24. Ratcliffe, D. A. *A Nature Conservation Review* Vol. **2**, 237 (Cambridge University Press, 1977).
25. Rycroft, D. W. thesis, Univ. Dundee (1971).
26. Mott, P. J. thesis, Univ. Dundee (1973).
27. Ingram, H. A. P., Rycroft, D. W. & Williams, D. J. A. *J. Hydrol. Amsterdam* **22**, 213–218 (1974).
28. Rycroft, D. W., Williams, D. J. A. & Ingram, H. A. P. *J. Ecol.* **63**, 557–568 (1975).
29. Rycroft, D. W., Williams, D. J. A. & Ingram, H. A. P. *J. Ecol.* **63**, 540–542 (1975).
30. Bormann, F. H. & Likens, G. E. *Pattern and Process in a Forested Ecosystem* (Springer, New York, 1979).
31. Weber, C. A. *Bot. Jb., Beibl.* **90**, 22 (1908).
32. Baden, W. & Eggelsmann, R. *Z. KultTech. Flurberein.* **4**, 240 (1963).
33. Granlund, E. *Sver. geol. Unders. Afh.* C(373) 60–70, 174–175 (1932).
34. Eggelsmann, R. & Schuch, M. in *Moor- und Torfkunde* (ed. Göttlich, K.) 153–154 (Schweizerbart'sche, Stuttgart, 1976).
35. Aartolahti, T. *Fennia* **93**, 31–32, 47–49, 169–170, 177–178 (1965).
36. Aaby, B. *Nature* **263**, 281–284 (1976).

Acumentin, a protein in macrophages which caps the 'pointed' end of actin filaments

Frederick S. Southwick & John H. Hartwig

Hematology–Oncology and Infectious Disease Units, Massachusetts General Hospital, Department of Medicine, Harvard Medical School, Boston, Massachusetts 02114, USA

Macrophages have a calcium-insensitive protein which binds to and inhibits monomer addition at the 'pointed' end of actin filaments labelled with heavy meromyosin. This protein, in concert with calcium-sensitive proteins already shown to bind the 'barbed' end of filaments, can lead to complete control of actin filament length in the cytoplasm. The blocking of actin filament ends in low as well as high calcium concentrations would prevent treadmilling in living cells.

HUXLEY[1] first recognized a polarity in actin filaments, and showed that heavy meromyosin bound to actin filaments, giving the appearance of arrowheads attached to the filaments. The direction of the arrowheads defines a 'pointed' and a 'barbed' end of the filaments, and during the contraction of striated muscle, actin moves in the direction of the pointed end. Subsequently, several investigators demonstrated that the association constant for monomer addition is greater at the barbed end of actin filaments decorated with heavy meromyosin than at the pointed end[2–4]. The barbed and pointed ends of actin filaments labelled with heavy meromyosin have therefore been called the preferred and unpreferred assembly ends, respectively. An observed inequality in dissociation to association rate constant ratio at the opposite ends of filaments at steady state[4] has given weight to the concept called treadmilling[5], which refers to a process in which actin monomers in equilibrium with actin polymers cycle within the filaments from the ends having the lower to the ends having the higher critical concentration. If such treadmilling exists it could cause rearrangement of actin-based structures in the cytoplasm of non-muscle cells[6]. The concept of treadmilling also predicts that anything which prevents an exchange of monomers with the barbed filament end would shorten the filament, and such capping at the pointed end would cause filament elongation.

Another manifestation of actin filament polarity is the binding of different molecules to one end of the filament. Recent work has identified several molecules that bind to the barbed end of actin filaments: these include gelsolin[7–10], a protein found in a wide variety of mammalian cells[11]; villin, a protein very similar to gelsolin and present only in the microvilli of intestinal and

0028-0836/82/210303—05$01.00

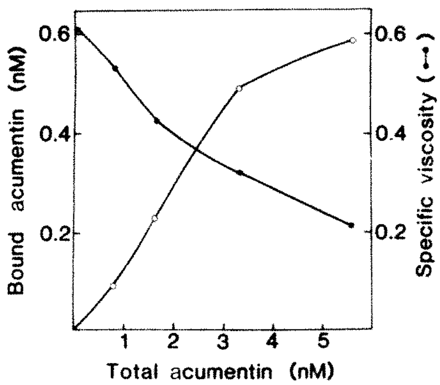


Fig. 1 Binding of acumentin to F-actin and correlation of its binding with changes in the final viscosity of actin. Rabbit skeletal muscle actin[30] 0.65 mg ml^{-1}, in 0.8 mM EGTA, 1.25 mM dithiothreitol, 0.05 mM $MgCl_2$, 0.5 mM ATP and 5 mM imidazole-HCl, *p*H 7.5, was gel-filtered on Sephadex G-150 and polymerized in the presence of increasing concentrations of acumentin by the addition of 3 M KCl to a final concentration of 0.1 M at 25 °C for 2 h. The specific viscosity of actin was determined in Cannon–Manning capillary viscometers and was 0.61 in the absence of acumentin. The outflow time for water in these viscometers was 48–50 s. The binding of acumentin to G-actin was determined by centrifuging 175 μl of the various actin solutions at 30 p.s.i. in a Beckman airfuge for 1 h to pellet all F-actin present. The amounts of acumentin in the pellet and supernatant fractions were determined by densitometry of Coomassie blue-stained polyacrylamide gels after electrophoresis in the presence of SDS. Acumentin alone did not sediment. Macrophage acumentin migrates as a single polypeptide band of molecular weight 65,000 (Tatsumi *et al.*, manuscript in preparation). This molecular weight was used to calculate the molar concentration of bound acumentin (○).

renal brush border epithelial cells[12–14]; a protein of *Acanthamoeba*[15]; and the cytochalasins[4,16,17]. Such molecules, in addition to binding to the barbed end of F-actin, stimulate nucleation of actin assembly and shorten the filaments whether added to actin during or after polymerization. The activities of gelsolin and villin are modulated by calcium. Gelsolin and villin bind to actin much more efficiently in the presence of micromolar calcium concentrations than in lower concentrations. The calcium-dependent shortening of actin filaments by gelsolin and villin provides a mechanism by which changes in calcium levels can regulate the length of actin filaments. However, to maintain total control over the filament length, it would be advantageous to cap both ends of the filaments because otherwise some filament annealing and monomer exchange might occur.

Table 1 Effect of macrophage acumentin and gelsolin on the direction of filament growth from F-actin decorated with heavy meromyosin

Type	—◄◄	◄◄—	-◄◄-	◄◄	—	(*n*)
F-actin alone	2.6 (2)	13 (10)	52 (40)	7.8 (6)	25 (29)	77
F-actin + acumentin	3.0 (2)	38 (26)	4.5 (3)	16 (11)	37 (25)	67
F-actin + gelsolin	20 (15)	1.3 (1)	1.3 (1)	52 (40)	26 (20)	77
F-actin + gelsolin + acumentin	2.0	—	—	84 (88)	24 (25)	102

Actin[30], 1 mg ml^{-1}, in 0.2 mM, $CaCl_2$, 0.5 mM 2-mercaptoethanol, 2 mM Tris-HCl *p*H 7.5, was polymerized in the absence or presence of gelsolin (molar ratio of 1 gelsolin to 50 actins) by the addition of 0.1 M KCl for 1 h at 25 °C and then incubated with or without acumentin (1 mol acumentin to 25 actins) for an additional 30 min at 25 °C. The filaments formed were then incubated with 1 mg ml^{-1} skeletal muscle heavy meromyosin for 10 min at 25 °C. 25 μm ml^{-1} of each was then mixed with 100 μg ml^{-1} of G-actin and the KCl concentration adjusted to 0.1 M and incubated at 25 °C for 5 min. The protein solutions were then placed in microtitre wells, and a strip of mica previously coated with carbon inserted such that the carbon layer floated on the surface of the solution but was not freed from the mica strip. The mica was then removed, allowing the carbon to re-attach, and inserted in a 1% uranyl acetate solution. The carbon was freed from the mica, Formvar–copper grids were dropped on to the carbon, and the grids were removed from the uranyl acetate solution with a strip of parafilm. The dried grids were photographed in a Philips 301 electron microscope and the negatives enlarged to a final magnification of ×60,000. Data are expressed in per cent with the number of filaments observed in parentheses. —◄◄, Growth of actin from the pointed end only, ◄◄—, barbed end only, -◄◄-, both ends, or ◄◄, lack of assembly onto heavy meromyosin-labelled nuclei, and —, filaments that did not have labelled nuclei attached.

β-Actinin, a protein isolated from skeletal muscle, is the only agent thus far suggested to bind the pointed end of actin filaments[18,19]. This protein inhibits the annealing of actin filament fragments. When annealing of a mixture of sonic fragments of actin decorated with heavy meromyosin and undecorated fragments occurred in the presence of β-actinin, Maruyama *et al.* had the impression that, compared with controls, more bare filaments attached to the barbed end of decorated segments than to the pointed end[19]. However, the barbed end of the decorated filament is in any case preferred for assembly, and therefore it is difficult to show an increase in assembly or annealing from that end. The existence of agents that block the barbed end of filaments makes it possible to ascertain blocking of the pointed end with more certainty. Using this approach we have found that a protein purified from human granulocytes[20] and rabbit lung macrophages binds to the pointed end of actin filaments. We propose to name this protein acumentin, from the Latin *acumen* meaning the point.

Acumentin binds and caps the pointed ends of actin filaments

Binding of acumentin to the pointed or unpreferred end of heavy meromyosin-labelled actin filaments was demonstrated indirectly in the electron microscope by observing the assembly of actin nucleated by actin filaments decorated with heavy meromyosin[2,3] in the presence and absence of acumentin. As shown in Table 1, the addition of decorated actin filaments to a solution of monomeric actin resulted in the growth of filaments from both ends of the nuclei, an observation in agreement with previous findings[2,3]. In some filaments growth was limited to only one end of the nuclei, with a bias of 5 to 1 favouring the barbed ends of the actin nuclei, the end known to have the most rapid assembly rate[4]. The addition of acumentin to actin nuclei markedly decreased the instances of actin assembly from the pointed ends of the nuclei. Although the findings are consistent with binding and capping of the pointed end of a filament by acumentin, this could also have been due to a second mechanism. The presence of unbound acumentin in the system could potentiate the formation of actin nuclei. Rapid filament assembly from these acumentin would decrease the actin monomer concentration, a condition that would favour the assembly of actin from the barbed or preferred assembly end of actin filaments decorated with heavy meromyosin. As demonstrated in Table 1, acumentin did increase the number of bare filaments.

Experiments using gelsolin ruled out this alternative. The gelsolin–calcium complex prevents assembly at the barbed end[8] and increases the actin monomer concentration, a condition which favours growth at the pointed end (Table 1). Addition

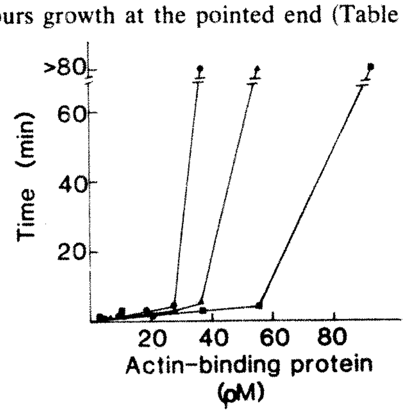


Fig. 2 Effect of increasing concentrations of acumentin on the critical concentration of macrophage actin-binding protein required for incipient gelation of F-actin. Actin, 5 mg ml^{-1}, in buffer A was polymerized by the addition of 3 M KCl to a final concentration of 0.1 M for 1.5 h at 25 °C. The F-actin was then diluted to 1 mg ml^{-1} in the presence or absence of acumentin and increasing concentrations of macrophage actin-binding protein. The solutions were mechanically mixed before being drawn up into capillary tubes. The final concentrations of acumentin were 0 (●), 20 (▲) and 30 (■) μm ml^{-1}, respectively. Gel points were determined as described by Yin *et al.* using a rolling ball viscometer[23].

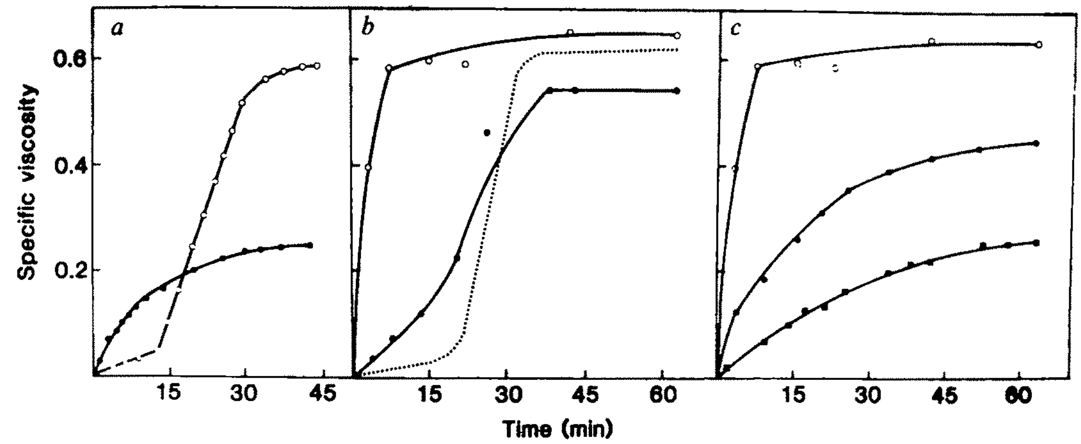


Fig. 3 Effects of acumentin on the kinetics of actin polymerization as determined by viscometry. *a*, Effect of acumentin on the nucleation of actin. Actin, gel-filtered on Sephadex G-150, was polymerized at a concentration of 0.65 mg ml^{-1} in 1.5 mM dithiothreitol, 0.175 mM ATP, 0.1 mM EGTA, 0.02 mM $MgCl_2$ and 5 mM imidazole-HCl, *p*H 7.5, in the presence (●) or absence (○) of 0.122 mg ml^{-1} of acumentin by the addition of 3 M KCl to a final concentration of 30 mM. *b*, Effect of acumentin on actin filament elongation. Gel-filtered actin, 0.5 mg ml^{-1}, in buffer A was polymerized by the addition of 1 M $MgCl_2$ to a final concentration of 1 mM in the presence (●) or absence (○) of 0.55 mg ml^{-1} of acumentin for 1 h at 25 °C. The filaments formed were used to nucleate actin assembly; 33 µl of each was then added to 300 µl of 0.5 mg ml^{-1} of G-actin in buffer A. The solution was immediately made up to 20 mM with 3 M KCl and 0.4 mM with 2 M $MgCl_2$ and its viscosity followed with time at 25 °C. The dotted line shows the polymerization of 0.5 mg ml^{-1} of actin in the absence of added acumentin F-actin nuclei. *c*, Additive effects of acumentin and gelsolin on actin filament elongation. 0.5 mg ml^{-1} of actin in the buffer detailed in Fig. 4*b* legend was polymerized in the presence of 33 µg ml^{-1} of gelsolin by the addition of 2 M $MgCl_2$ to a final concentration of 1 mM. The gelsolin–F-actin complexes were then incubated with or without acumentin, 0.55 mg ml^{-1}, for 1 h. 50 µg ml^{-1} of F-actin alone (○), gelsolin–F-actin (●) or gelsolin–F-actin–acumentin solution (■) were added to 0.5 mg ml^{-1} of G-actin. The solutions were immediately made 20 mM and 0.4 mM in KCl and MgCl (ref. 2) by addition of 3 M KCl and 2 M $MgCl_2$, respectively.

of acumentin to the preparation of labelled nuclei almost completely abolished growth from the labelled filaments, and there was no increase in the number of unlabelled filaments. The data establish that acumentin caps the pointed end. Experiments described below using biophysical techniques are all consistent with the conclusion that acumentin binds to the ends of actin filaments.

Assembly of actin with acumentin shortens the filament length distribution

When actin is polymerized in the presence of acumentin, its final viscosity is decreased relative to that of actin assembled in its absence[20]. Figure 1 shows the correlation between the capacity of acumentin to bind to actin during polymerization with its ability to reduce the final equilibrium viscosity of actin after assembly. The binding of acumentin to actin filaments increased in proportion to the decrease in final viscosity of the actin solution. This effect of acumentin on the viscosity of actin resulted from the formation of shortened actin filaments, because acumentin does not increase the fraction of monomeric actin in equilibrium with actin filaments. Although the ability of acumentin to reduce the final viscosity of actin was independent of calcium concentration (10^{-8}–10^{-4} M), its effects on actin viscosity could be inhibited by KCl concentrations above 0.1 M, complete inhibition occurring at 0.3 M KCl (ref. 20). Binding was determined by co-sedimenting acumentin bound to filaments polymerized in its presence. In this assay, the amount of acumentin bound to actin correlated with the degree of filament shortening, and binding increased in almost direct proportion to the concentration of acumentin added. At any given acumentin concentration only 12.6±2% (mean±s.d.) of the added acumentin was bound. As acumentin was present during actin assembly, each molecule theoretically could have bound monomers to form nuclei, suggesting that low affinity rather than a limited capacity was responsible for this low level of binding. The observed linear increase in binding is consistent with the conclusion that acumentin stimulates actin nucleation (Fig. 3*a*), thereby increasing filament number. Therefore, the amount of acumentin bound will be directly related to the number of filament ends.

The conclusion that filaments were shortened is further supported by the finding that acumentin increased, in a concentration-dependent fashion (Fig. 2), the critical amount of cross-linker required to gel a solution of actin. As the critical gel point is inversely related to filament length[21], acumentin must decrease the average length of actin filaments. The weight-average lengths of actin filaments polymerized in the presence of 20 and 30 µg ml^{-1} of acumentin can be calculated directly from the gel point determinations, using the Flory equation: $V_c = C/X_w$, where V_c is the critical gel point in mol^{-1}, C is the polymer concentration in g l^{-1} and X_w is the weight-average molecular weight of the actin filaments[21]. The calculated length of actin filaments polymerized in the absence of acumentin was 2.44 µm, and 1.78 and 1.2 µm in the presence of 20 and 30 µg ml^{-1} of acumentin, respectively.

A comparison of the gel point values with the binding data supports the idea that one acumentin molecule is bound per actin filament. Assuming that 12.6% of the added acumentin was bound to actin filaments, then for 1.0 mg ml^{-1} of actin, 1 mol of acumentin would be bound for every 609 and 409 actin monomers in filaments for 20 and 30 µg ml^{-1} of acumentin added, respectively. Dividing these values by the known number of monomers per unit filament length (370 monomers per µm (ref. 22)) gives respective lengths of 1.64 and 1.11 µm. The values are in reasonable agreement with those determined from the gel point values.

We studied the effect of acumentin on the nucleation and elongation steps of actin polymerization and on filament annealing to analyse the mechanism by which acumentin affects filament length. As shown in Fig. 3*a*, acumentin accelerates the onset of actin polymerization; this suggests that acumentin promotes the nucleation of actin monomers. Growth of filaments from these nuclei may result in the formation of a larger number of shortened filaments relative to actin assembled by spontaneous nucleation, because more nuclei are competing for the monomer pool[8]. However, this effect alone is insufficient to maintain the filaments in a shortened state because such filaments would reanneal with time to a normal length distribution. Short filaments are stabilized at equilibrium by the fact that acumentin binds to their ends (Fig. 3*b*) and prevents annealing and redistributing to a longer length distribution, as amplified below.

To study the elongation phase of filament assembly, nucleation was stimulated maximally by adding nuclei in the form of actin filaments to solutions of monomeric actin. The addition of the F-actin nuclei to monomeric actin resulted in a rapid polymerization of actin, as judged by an immediate increase in the viscosity of the solution (Fig. 3*b*). On the other hand, the addition to monomeric actin of nuclei in the form of actin filaments which had been assembled in the presence of acumentin, promoted a slower initial rate of assembly than did the addition of actin nuclei alone. Therefore, nuclei capped with acumentin, like nuclei reacted with other agents which bind to the end of actin filaments[8,10,14–16], are less effective than

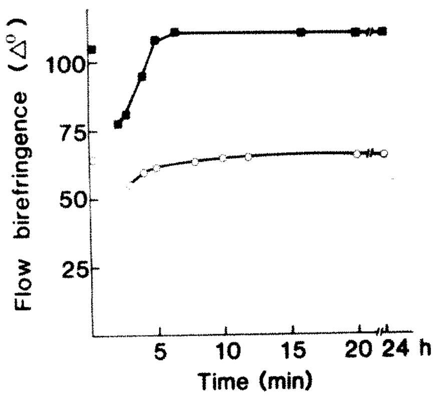


Fig. 4 Effect of acumentin on the annealing of F-actin fragments. F-actin, 0.25 mg ml^{-1}, in 0.1 M KCl, 0.2 mM $CaCl_2$, 0.5 mM ATP, 0.5 mM 2-mercaptoethanol, 2 mM Tris-HCl, *p*H 7.5, was mixed on a mechanical mixer with (■) or without (○) 32 μg ml^{-1} of acumentin. The flow birefringence and extinction angles of the two solutions were determined as described by Maruyama[31]. The solutions were then sonicated for 40 s on ice and the recovery of birefringence from sonication was then followed with time. F-actin had a final extinction angle of 3°. Acumentin increased the extinction angle to 8°. Velocity gradient was 100 s^{-1}.

actin nuclei in stimulating the growth of actin. This effect is a consequence of the decrease in the number of ends on the nuclei available for binding of actin monomers.

Acumentin and gelsolin have additive inhibitory effects on filament elongation (Fig. 3*c*). When actin filament fragments capped with gelsolin were added to G-actin solutions, the onset of actin assembly was immediate as judged by the increase in the viscosity of the solution. However, the initial rate of filament assembly was lower than that of polymerization nucleated by addition of actin filaments alone. The addition of acumentin to the nuclei capped by gelsolin further depressed the rate of filament assembly without affecting the onset of polymerization. The combination of the two proteins might at first glance be expected to block both ends of the nuclei, thereby causing actin to assemble as if no nuclei were added. There are two reasons this result was not observed. First, gelsolin affects actin polymerization in several ways. It reduces the number of free nuclei ends by binding to the barbed end of actin filament nuclei. However, gelsolin also 'severs' actin filaments, thereby creating new pointed ends[23,24]. Moreover, free gelsolin nucleates actin monomers. Second, acumentin also reduces the number of free filament ends by binding to the pointed end of the filaments. However, free acumentin can also nucleate actin monomers. Therefore, as free molecules of acumentin and gelsolin were not removed in the conditions of the experiment shown in Fig. 3*c*, unbound gelsolin and acumentin would be present in the solution. These molecules could nucleate actin monomers, producing new filaments with one free end. In addition, gelsolin would be expected to produce new ends by cutting filaments. Therefore, the combined effect of gelsolin and acumentin is explained most simply by a reduction in, but not a complete removal of, filament ends.

As shown in Fig. 4, acumentin inhibits the annealing of broken actin filaments. When actin was disrupted by sonication in the absence of acumentin, flow birefringence values recovered rapidly. Shearing of actin filaments in the presence of acumentin on a mechanical mixer or sonication caused an immediate and persistent decrease in the flow birefringence value of the actin, indicative of filament shortening. The decrease in birefringence after shearing is consistent with binding of acumentin to the ends of actin filaments which had been fragmented by shear and with the inhibition of annealing of these capped fragments.

Biological implications

Acumentin binds and caps the pointed end of actin filaments, and potentiates the nucleation step of actin assembly. This effect increases the number of nuclei available for addition of actin monomers and distributes the resulting polymer mass over more filaments, thus shortening the filament length distribution. The filaments elongate more slowly and remain short because acumentin caps one of their ends. Whether acumentin shortens filaments by actively severing them or by inhibiting broken filaments from annealing is unknown. All experiments involved mechanical shearing of the filaments, and transient filament fragmentation invariably results from these manipulations[25].

Phagocytic leukocytes have a layer of distinct cytoplasm located just beneath their plasma membrane, and this cortical cytoplasm is redistributed during cell movement. Electron micrographs of the cortex of these cells reveal the presence of short actin filaments 100–300 nm long[26–29].

The abundance of acumentin in leukocytes may in part explain the existence of these short filaments. The molar ratio of acumentin to actin in cytoplasmic extracts is 1 to 5 (ref. 20). Assuming that 10% of the acumentin in the cortex were bound to actin, 1 acumentin molecule would cap a filament containing 50 actin molecules or one which was 135 nm long. We presume that acumentin would be constitutively bound to actin filaments, the amount bound being determined solely by its affinity for actin, because its binding to actin is equivalent in high and low calcium concentrations. The gelsolin–calcium complex confers the remaining control and the fine regulation of filament length.

Exchange of actin monomers with the ends of filaments would be totally inhibited if both ends were capped by acumentin and gelsolin, a condition that would occur in the cell when the free calcium level is greater than micromolar. When the free calcium concentration decreases below micromolar concentrations gelsolin dissociates from the barbed end of the filament. Filaments with freed barbed ends would compete with acumentin for the pointed ends of the filaments. Our binding data suggest that the affinity of acumentin for the pointed end of actin filaments is lower than that of actin monomers for the barbed end of filaments[4], and probably lower than the affinity of filament ends for each other during annealing. High concentrations of actin filaments with free barbed ends could displace acumentin from the pointed ends of other filaments. In these conditions the average filament length would increase.

When the free calcium concentration rises sufficiently, gelsolin would split actin filaments and remain bound to their barbed ends. This would decrease the number of barbed and increase the number of pointed filament ends. In the absence of acumentin, residual free barbed filament ends would anneal with the pointed ends of other filaments and free monomers could also bind to the pointed ends. Capping of the pointed ends by acumentin would prevent either this filament annealing or monomer addition, thereby conferring a greater degree of precision to the control of actin filament length by the calcium–gelsolin complex.

Furthermore, the interaction of acumentin and gelsolin with actin could permit rapid changes in legnths of existing filaments without concomitant alterations in a pool of monomeric actin which might cause the uncontrolled formation of new filaments. Finally, the presence of an agent that constitutively blocks a filament end, renders unlikely the existence of actin treadmilling in the living cell.

This work was supported by USPHS grants HL19429, CA 09321, R23 AI 18294 and by the Muscular Dystrophy Association, the Edwin S. Webster foundation and E. W. Hiam. We thank Drs Thomas P. Stossel and Ken S. Zaner for help in determining the mechanism of action of acumentin, Dr Helen Yin for gelsolin and Mr Richard Noone Jr for his assistance in the selection of the name of acumentin.

Received 8 January; accepted 18 March 1982.

1. Huxley, H. E. *J. molec. Biol.* **3**, 281–308 (1963).
2. Woodrum, D. T., Rich, S. A. & Pollard, T. D. *J. Cell Biol.* **67**, 231–237 (1975).
3. Kondo, H. & Ishiwata, S. *J. Biochem.* **79**, 159–171 (1976).
4. Pollard, T. D. & Mooseker, M. S. *J. Cell Biol.* **88**, 654–659 (1981).
5. Wegner, A. *J. molec. Biol.* **108**, 139–150 (1976).
6. Kirschner, M. W. *J. Cell Biol.* **86**, 330–334 (1980).

7. Yin, H. L. & Stossel, T. P. *Nature* **281**, 583–586 (1979).
8. Yin, H. L., Hartwig, J. H., Maruyama, K. & Stossel, T. P. *J. biol. Chem.* **256**, 9693–9697 (1981).
9. Lind, S., Yin, H. L. & Stossel, T. P. *J. clin. Invest.* (in the press).
10. Wang, L. L. & Bryan, J. *Cell* **25**, 637–649 (1981).
11. Yin, H. L., Albrecht, J. H. & Fattoum, A. *J. Cell Biol.* **91**, 901–906 (1981).
12. Bretscher, A. & Weber, K. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2321–2325 (1979).
13. Bretscher, A. & Weber, K. *Cell* **20**, 839–847 (1980).
14. Glenney, J. R., Kaulfus, P. & Weber, K. *Cell* **24**, 471–480 (1981).
15. Isenberg, G., Aebi, U. & Pollard, T. *Nature* **288**, 455–459 (1980).
16. MacLean-Fletcher, S. & Pollard, T. D. *Cell* **20**, 329–341 (1980).
17. Brenner, S. L. & Korn, E. D. *J. biol. Chem.* **254**, 9982–9985 (1979).
18. Maruyama, K. *J. Biochem.* **69**, 369–386 (1971).
19. Maruyama, K. *J. Biochem.* **81**, 215–232 (1977).
20. Southwick, F. S. & Stossel, T. P. *J. biol. Chem.* **256**, 3030–3036 (1981).
21. Flory, P. J. *J. phys. Chem.* **46**, 132–140 (1942).
22. Hanson, J. & Lowy, J. *J. molec. Biol.* **6**, 46–60 (1963).
23. Yin, H. L. & Stossel, T. P. *J. biol. Chem.* **255**, 9490–9493 (1980).
24. Yin, H. L., Zaner, K. & Stossel, T. P. *J. biol. Chem.* **62**, 647–659 (1974).
25. Asakura, S., Taniguchi, M. & Oosawa, F. *J. molec. Biol.* **7**, 55–69 (1963).
26. Stossel, T. P. *Proc. R. Soc.* B (in the press).
27. Trotter, J. A. *Expl Cell Res.* **132**, 235–248 (1981).
28. Niederman, R. *Biophys. J.* **37**, 282a (1982).
29. Boyles, J. & Bainton, D. F. *J. Cell Biol.* **82**, 347–368 (1970).
30. Spudich, J. A. & Watt, S. *J. biol. Chem.* **246**, 4866–4871 (1971).
31. Maruyama, K. *J. Biochem.* **535**, 277–286 (1964).

Protamine is an inhibitor of angiogenesis

Stephanie Taylor & Judah Folkman*

Department of Surgery, Children's Hospital Medical Center and the Departments of Anatomy and Surgery, Harvard Medical School, Boston, Massachusetts 02115, USA

Protamine is shown to be a specific inhibitor of angiogenesis. The compound inhibits the capillary proliferation observed in embryogenesis, inflammation, certain immune reactions and the growth of solid tumours.

ANGIOGENESIS, the growth of new capillary blood vessels, is important in normal processes such as development of the embryo, formation of the corpus luteum and wound healing. It is also a component in pathological processes such as chronic inflammation, certain immune responses and neoplasia[1–3]. Furthermore, angiogenesis is a property of most solid tumours and is necessary for their continued growth[4]. Consequently, we have hypothesized that an angiogenesis inhibitor might be used to control tumour growth[5].

During our early studies of angiogenesis we observed that mast cells accumulated at a tumour site *before* the ingrowth of new capillary sprouts, but that mast cells alone could not initiate angiogenesis[6]. We later postulated that mast cells might facilitate the continuous migration[7] of capillary endothelial cells towards the solid tumour. When capillary endothelial cells were cloned and carried in long-term culture[8], and a quantitative assay was devised to measure their migration rate[9], we found that heparin released by mast cells increased the migration (chemokinesis) of capillary endothelial cells *in vitro*[10]. Protamine, an arginine-rich basic protein of 4,300 molecular weight (M_r), found only in sperm and known for its capacity to bind heparin, blocked the ability of mast cells and heparin to stimulate migration of capillary endothelial cells[10].

We now show that protamine is a specific inhibitor of stimulus. Protamine released locally prevents the neovascularization induced by an inflammatory agent or by an immune reaction and it inhibits capillary proliferation in the embryo. Protamine also inhibits tumour angiogenesis and subsequent tumour growth when it is applied locally, but in only a few cases when it is administered systemically. It has no effect on established capillaries that are not proliferating. We further show that heparin can enhance the intensity of angiogenesis induced by tumour extract *in vivo* but that heparin alone cannot initiate angiogenesis.

Inhibition of angiogenesis in the embryo

The effect of protamine on growing embryonic vessels was investigated using chick embryos of different ages. Capillaries appear in the yolk sac at 48 h and grow rapidly over the next 6–8 days. Protamine in methylcellulose disks applied locally inhibited capillary growth and produced a large avascular zone in the yolk sac (Fig. 1), whereas control embryos implanted with empty disks did not develop avascular zones.

Heparin added to the protamine disk prevented the development of the avascular zone (Fig. 1). The non-anticoagulant fraction of heparin[11] was equally effective in this action, while heparin alone did not cause an avascular zone.

In contrast, protamine had no effect on mature, non-growing[12] vessels studied on the chorioallantoic membrane at day 10. However, protamine applied to younger proliferating vessels of the day 5 or 6 chorioallantoic membrane produced a large avascular zone within 48 h.

Inhibition of tumour angiogenesis

The first indication that protamine could inhibit tumour angiogenesis came from studies of heparin on the chick chorioallantoic membrane (CAM). When 100 μg of tumour extract (partially purified for angiogenesis activity from human hepatoma cells)[9] was placed on the CAM, new capillary growth was elicited within 48 h in 90% of the eggs. Addition of heparin (6 μg, that is, 1 unit) reduced to 25 μg the amount of tumour extract required to elicit neovascularization in 100% of eggs, and reduced the interval between implantation and capillary growth to 24 h (Fig. 2). Heparin alone had no effect. Addition of protamine just sufficient to neutralize heparin (12 μg) eliminated the angiogenesis enhancing effect of heparin. However, when protamine was increased to 20 μg, it prevented all angiogenesis induced by tumour extract (25 μg) (Fig. 2).

A further test used a sustained release pellet[13] of protamine positioned between tumour and normal vessels in the rabbit cornea[14], allowing quantification of the inhibition of new capillary proliferation as a function of the rate of linear capillary growth. A protamine pellet juxtaposed to a V2 carcinoma implant almost completely inhibited capillary growth. In contrast, in the untreated eyes, tumours were vascularized by hundreds of new capillary sprouts and grew rapidly. Tumours in the protamine treated eyes did not become vascularized until the protamine pellet was removed. The mean vessel growth rate in the protamine-treated eyes was 0.02 ± 0.01 mm per day compared with 0.25 ± 0.07 mm per day in the untreated eyes (Figs. 3*a*, *b* and 4).

A more rigorous test of an angiogenesis inhibitor is to implant tumour directly into vascular tissue. Thus when V2 carcinoma and protamine pellets were implanted in a subcutaneous pouch[15] in the rabbit ear, the density of neovascularization (observed by transilluminating the ear) was markedly reduced in comparison with tumour in the opposite ear not exposed to protamine. Although vessel growth rates could not be quantified, tumour growth could be. Protamine-treated tumours grew at a mean rate of 8 ± 1 mm^3 per day compared with 68 ± 5 mm^3 per day for untreated tumours (Fig. 5). By histological section, the protamine-treated tumours, while still

* To whom correspondence should be addressed: Children's Hospital Medical Center, 300 Longwood Avenue, Boston, Massachusetts 02115, USA.

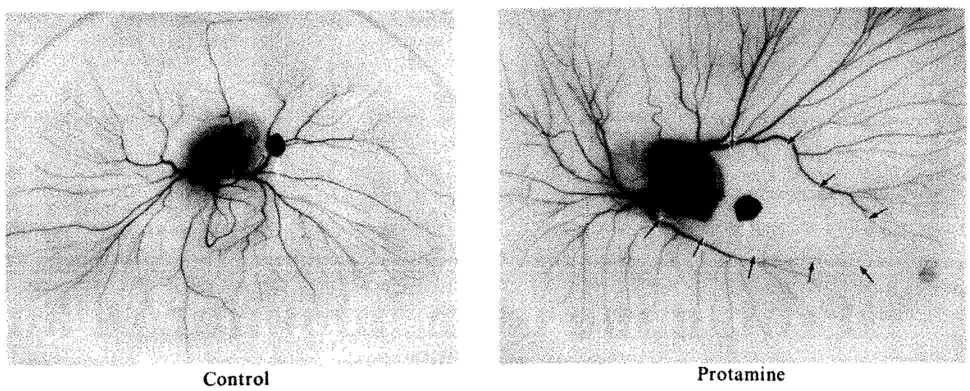


Fig. 1 Inhibition of angiogenesis by protamine in the chick embryo yolk-sac vascular membrane. Fertilized embryos are placed in a Petri dish on day 4 and cultured in a humidified incubator in 5% CO_2 as previously described[32]. Protamine sulphate (Calbiochem) is dissolved in distilled H_2O to make 14.3 $\mu g\ \mu l^{-1}$. 500 μl of this solution is added to 500 μl of 1% (w/v) methylcellulose (in distilled water, previously sterilized by autoclaving). Aliquots (7 μl) of this mixture are pipetted onto Teflon moulds of 3.2 mm diameter and allowed to dry (J. Barlow and J.F., unpublished). The resulting methylcellulose disk contains 50 μg protamine and can be peeled away from the Teflon mould. For control disks, 500 μl H_2O is substituted for the protamine. For photographic purposes, India ink can be added to the methylcellulose (2 μl per 100 μl). A methylcellulose disk (white arrow) was placed at the advancing edge of the vascular membrane on day 4 and the resultant avascular zone (black arrows) measured on day 6. The outlines of the yolk sac and of the avascular zone were traced on top of the Petri dish cover and converted to area with a Zeiss MOP-3 digital image analyser (Carl Zeiss). Forty-three embryos were implanted with protamine disks. The mean area of the vascular membrane on day 6 was $24.8 \pm 1.0\ cm^2$. The mean area of the avascular zones produced by protamine was $3.0 \pm 0.4\ cm^2$, or 12.5% of the original vascular area. In 43 control embryos implanted with empty disks, the mean area of the vascular membrane was $23.6 \pm 1.0\ cm^2$. There was no avascular zone in these embryos. Heparin (25 μg) added to the protamine disk prevented the development of the avascular zone. The non-anticoagulant fraction of heparin[11] was equally effective in preventing the protamine-avascular zone. Heparin alone did not cause an avascular zone. While the size of the avascular zone was dose related to protamine, it was not specific for protamine. Extremes of pH and osmolarity could also produce an avascular zone.

viable, were either avascular or were penetrated only sparsely by capillaries. In contrast, the untreated tumours were pervaded by a dense capillary network.

Inhibition of inflammatory angiogenesis

To determine whether protamine could inhibit inflammatory angiogenesis[16], a focus of chronic inflammation was established in the rabbit cornea by implanting silica particles. In each of six rabbits, a protamine pellet (prepared as described in Fig. 3 legend) was implanted between the silica and the normal vascular bed. An empty pellet was implanted in the opposite eye. New capillaries appeared at 2 days in the untreated eyes and grew towards the silica at a mean rate of 0.2 ± 0.02 mm per day. By contrast, in the protamine-treated eyes the mean vessel growth rate was 0.06 ± 0.01 mm per day. When the protamine pellet was exhausted or removed, new capillaries advanced towards the silica at the same rate as in untreated eyes.

Inhibition of immune angiogenesis

To determine whether protamine could block immune angiogenesis[17], fragments of rabbit lymph node were implanted in corneal pockets of nine outbred rabbits. Either a protamine pellet or an empty pellet was implanted between the node and the normal vascular bed. In the absence of protamine, new capillaries entered the cornea by day 4 and advanced towards the lymph node at a mean rate of 0.20 ± 0.03 mm per day. In the presence of protamine, angiogenesis was suppressed completely until day 8, when a few vessels began to grow at 0.05 ± 0.03 mm per day.

Systemic administration of protamine

Although the systemic administration of protamine was limited by its toxicity at high doses (lethargy, weakness and occasionally sudden death), it has, within the tolerated doses, been used to test the inhibition of growth of four types of lung metastases and three of their subcutaneous primary tumours.

Protamine concentrations in the lung were fivefold greater than in the subcutaneous tissue, and 40-fold greater than in the skin.

Lung metastases. To test whether systemically administered protamine was an effective inhibitor of tumour growth, mice with lung metastaases were treated with subcutaneous protamine at 12-h intervals (Fig. 6). When mice which had received Lewis lung tumour cells, B16 melanoma cells (subcutaneous strain) or B16 melanoma (tissue culture strain) either intravenously (i.v.) or subcutaneously (s.c.) were treated with a well tolerated dose of protamine (60 mg per kg, s.c., every 12 h), there was 77–92% inhibition of mean lung metastases tumour volume compared with saline-treated animals. The number of lung metastases did not differ significantly between protamine- and saline-treated animals with the exception of the two strains of melanoma.

Similarly, when rats which had received Walker 256 carcinoma cells i.v. were treated with a well tolerated dose of protamine (120 mg per kg, s.c., every 12 h), there was 97% inhibition of mean lung metastases tumour volume compared with saline-treated animals (Fig. 6). The number of metastases was also reduced: there were 3 metastases per rat in the protamine-treated animals compared with 64 metastases per rat in the saline-treated animals (Fig. 6).

Subcutaneous tumours. The growth of neither Lewis lung carcinomas nor B16 melanoma (subcutaneous strain) when implanted s.c. was significantly inhibited by protamine except for some inhibition of the former at highly toxic doses (135 mg per kg). However, systemic protamine did inhibit the growth of subcutaneous tumours of B16 melanoma (tissue culture strain), as will now be described.

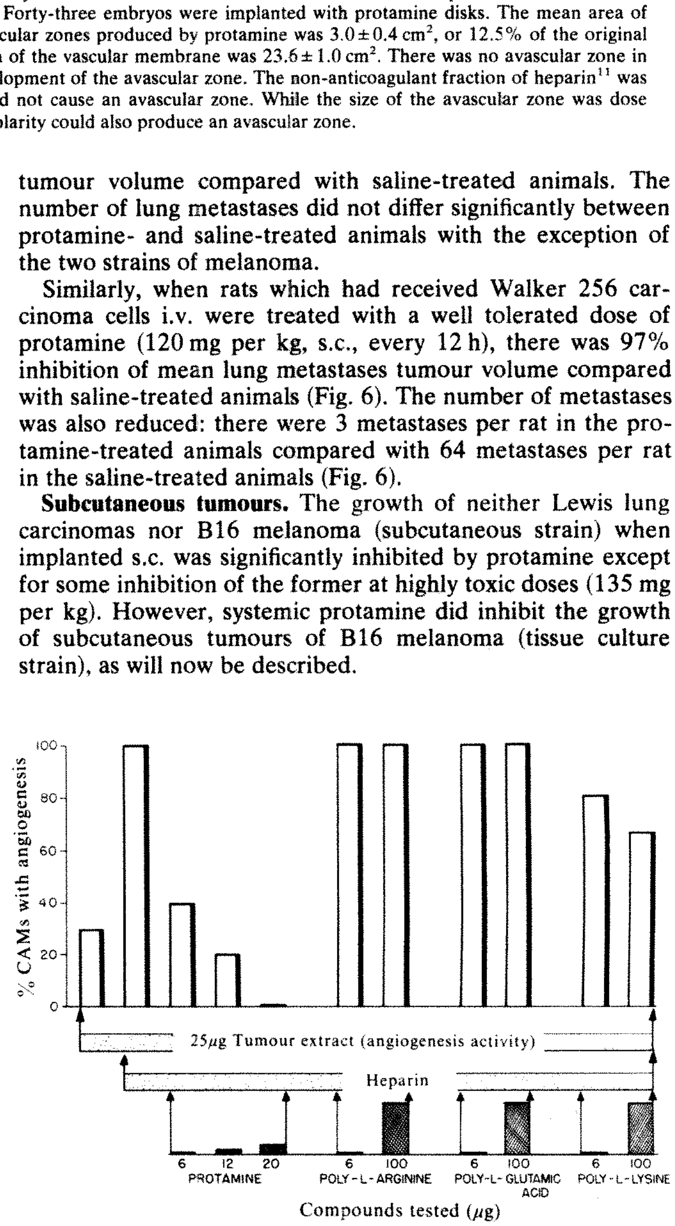


Fig. 2 Angiogenesis on the chick chorioallantoic membrane (CAM). Tumour extract (25 μg, from hepatoma cells[9]) in 5 μl H_2O was placed on the centre of a round plastic coverslip (Thermanox, 15 mm diameter) (J. Barlow and J.F., unpublished data). To each disk was added another 5 μl aliquot containing either heparin (6 μg) or one of the test compounds. The coverslip was allowed to dry, placed on the CAM of 9-day-old embryos and viewed 24 h later with a $\times 12$ stereomicroscope. Angiogenesis was present if new capillaries converged on the spot in the centre of the coverslip. Controls were poly-L-arginine (M_r 130,000, Sigma), poly-L-glutamic acid (M_r 10,000) and poly-L-lysine (M_r 3,000). The five bars on the left represent 50 eggs each. The remaining six bars to the right represent 15 eggs each.

Fifty-two C57BL/6 mice received 4.6×10^7 melanoma tissue culture strain F2[18] s.c. (smaller inocula of 10^3, 10^4 and 10^5 cells also produced tumours, although much later). Half of the mice received protamine 60 mg per kg, s.c., every 12 h in a remote site and the other half received saline.. All tumours took and grew at approximately the same rate in both protamine- and saline-treated animals until day 5 when the tumours were nearly 0.05 cm^3 and had become vascularized. The protamine-treated tumours then slowed their growth (0.05 cm^3 per day) until day 18 when they either stopped growing or began to regress. The saline-treated tumours continued to grow rapidly over the same period (0.23 cm^3 per day). In 13 (50%) of the protamine-treated mice, tumour regressed completely. It was necessary to lower the protamine dose by approximately one-half once regression was complete. These mice remained tumour free after protamine was discontinued. Eight (30%) died while the tumour was undergoing rapid necrosis. Two mice died with melanoma ascites after the primary tumour had disappeared. In three mice, tumours underwent incomplete regression and remained as small, barely visible lesions. Their tumour growth resumed whenever protamine was discontinued, and they were dead by day 68. The saline-treated animals died with massive tumours (17.8 ± 0.4 cm^3) by day 30. Using the same strain of melanoma, this experiment was repeated eight times over a 2 yr period with essentially similar results (tumour regression in ~50% of the animals). After a large tumour had regressed to become no longer visible, it was necessary to continue the lower dose of protamine for at least 2 more weeks, otherwise the tumour recurred. When mice that had been tumour free and not receiving protamine for 1–11 months received a second s.c. inoculation of the same strain of melanoma cells, new tumours developed and were also susceptible to protamine inhibition.

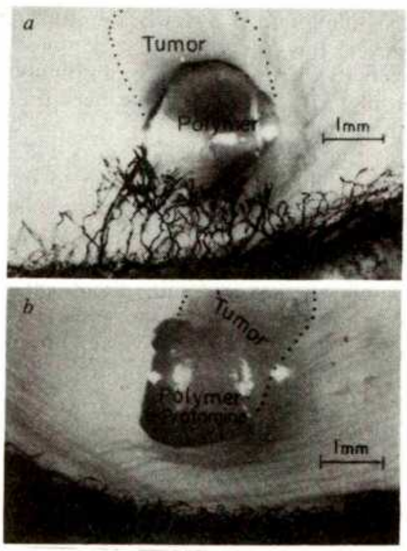


Fig. 3 Inhibition of tumour angiogenesis in the rabbit cornea by protamine. Polymer pellets of ethylene vinyl acetate (EVA) copolymer, of ~1 mm diameter were impregnated with 1,000 μg protamine sulphate as previously described[13]. The pellets were coated with 2% EVA in methylene chloride, dried for 3 h and sterilized by UV light for 2 min. The pellets were implanted in a pocket in the rabbit cornea, 1 mm from the limbus. A 1 mm^3 piece of V2 carcinoma was implanted distal to the polymer, 2 mm from the limbus[14]. In the opposite eye of each rabbit, control polymer pellets that were empty were similarly implanted in juxtaposition to the tumour. By spectrophotometric analysis *in vitro*, the pellets released 400 μg protamine on day 1. This sustained release declined exponentially until day 4, when the pellets released an average of 15 μg per day thereafter, for the next 15 days. As capillary blood vessels grew towards the tumour implant, maximum vessel length was measured every 3 days with a stereoscopic slit-lamp at ×10 (±0.1 mm). At day 10 a typical rabbit was killed and India ink was injected into each carotid artery. The corneas were removed and photographed with a stereoscope. *a*, New capillary blood vessels growing towards the tumour and passing over the blank polymer. *b*, In the presence of the protamine pellet (opposite eye), new capillary vessels at the edge of the cornea do not enter the cornea. Histological sections showed viable tumour in each cornea.

Thus, the first tumour regression apparently was not accompanied by a strong immune reaction.

Controls

To control for the specificity of protamine as an angiogenesis inhibitor, compounds similar in structure or charge density to protamine were tested against tumour angiogenesis in the CAM and against growing embryonic vessels on the yolk sac. Poly-L-arginine, poly-L-glutamic acid and poly-L-lysine produced no inhibition of angiogenesis at concentrations two to four times the protamine concentration that was inhibitory (Fig. 2). At five times the protamine concentration, poly-L-lysine weakly inhibited tumour angiogenesis in the CAM but did not inhibit the growth of yolk sac vessels. These compounds were all lethal when applied at concentrations greater than five times the protamine concentration. Platelet factor 4 and major basic protein from eosinophils have a high affinity for heparin. They both produced avascular zones at approximately one-half the concentration of protamine.

As another control of protamine specificity, mixtures of heparin and protamine (1:2) were assayed in the chick embryo (Fig. 1). Both the commercial anticoagulant heparin (Sigma) and the non-anticoagulant fraction[11] (courtesy of Dr Robert Rosenberg) neutralized the anti-angiogenesis activity of protamine in the yolk sac and on the CAM. Heparin (Sigma) was also injected into tumour-bearing mice simultaneously with protamine (but at a different site). Inhibition of tumour growth was abolished; tumour size and mouse mortality were the same as for mice injected with saline alone (data not shown).

To exclude the possibility that protamine inhibition of tumours might be caused by direct cytotoxicity, tumour cells were exposed to protamine *in vitro*. Protamine was neither cytostatic nor cytotoxic to V2 carcinoma, B16 melanoma (tissue culture strain), Lewis lung or Walker cells at concentrations of 50–500 μg ml^{-1}. Growth suppression began at 500 μg ml^{-1} for Walker cells, but melanoma, V2 and Lewis lung cells were not suppressed at 1,000 μg ml^{-1}. In fact, growth of V2 carcinoma was stimulated to 55% above control by a protamine concentration of 100 μg ml^{-1}, and Lewis lung cells were stimulated to 37% above control by a protamine concentration of

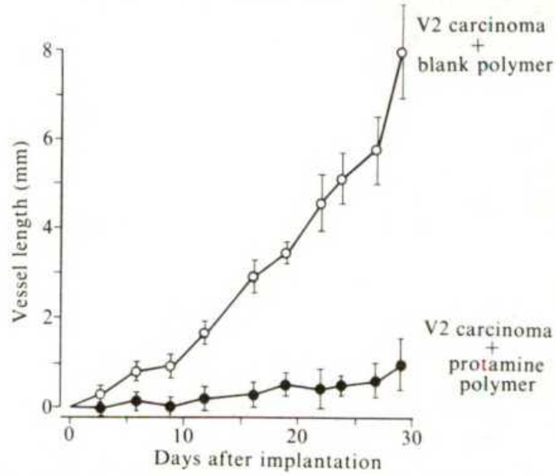


Fig. 4 Measurement of capillary growth rate in the rabbit cornea. Series of nine rabbits with V2 carcinoma in each cornea accompanied by a protamine pellet (●) or empty pellet (○) were used as described in Fig. 3 legend. In the presence of protamine, capillary growth is suppressed until the week 3 (when protamine release from the polymer pellet declines). In the eyes containing a blank pellet, tumours all became vascularized and grew to a large exophytic mass, necessitating the death of these animals by about 30 days. In a parallel experiment (data not shown), in which both eyes of an animal were treated with protamine, capillary growth resumed within 3 days of removal of the pellet. There was subsequent rapid tumour growth. Protamine pellets implanted in the absence of tumour did not induce neovascularization. Measurements are expressed as mean ±s.e.m. here and in all subsequent figure legends. Eyedrops of protamine in Ringer's solution (100 mg ml^{-1}) applied three times per day, did not inhibit capillary growth. The eyedrops caused no detectable irritation in the normal eyes of rabbits. Systemically administered protamine failed to inhibit capillary growth in the cornea.

500 μg ml^{-1}. These concentrations exceed any local concentrations of protamine achieved *in vivo* by sustained-release pellets. Furthermore, when protamine was administered systemically, its peak uptake in subcutaneous melanoma was 70 μg per 100 mg of tumour at 1 h. This is equivalent to 0.001 μg per 1,000 cells (assuming that a 100 mg tumour contains at least 10^7 cells). The melanoma cells *in vitro* grew maximally at 2 μg per 1,000 cells, that is, at more than 2,000 times the protamine concentration achieved *in vivo*.

To determine if protamine was cytotoxic to rapidly dividing cells *in vivo*, melanoma-bearing mice treated with protamine or saline for 11 days received 80 μCi ^{3}H-thymidine and were killed 5 h later. The percentage of ^{3}H-thymidine labelled cells in the bone marrow was similar for the protamine-treated (20%) and saline-treated mice (17%). In the saline-treated tumours, 17% of the tumour cells were labelled. There was no necrosis. In contrast, the regressing protamine-treated tumours were marked by a large necrotic centre and a few outer layers of viable cells. The labelling index of the rim of viable cells was 16%, not significantly different from the saline-treated tumour.

To exclude the possibility of cumulative toxicity at the 60 mg per kg dose, 10 non-tumour-bearing mice received protamine (60 mg per kg, s. c., every 12 h) for 1 yr, beginning at age 6 weeks. Their weight gain and activity were the same as for 10 saline-treated mice throughout this period, and all 20 mice were alive at the end of 1 yr.

Discussion

These studies show that protamine is an inhibitor of angiogenesis in the four species tested, irrespective of whether the angiogenic stimulus is embryological, neoplastic, inflammatory or immunogenic.

The mechanism by which protamine inhibits angiogenesis is unclear. However, previous *in vitro* studies suggest that heparin modulates the rate of migration of capillary endothelium[10]. The specific binding of protamine to heparin could block the linear migration of capillary endothelial cells, thus interfering with a crucial step in capillary formation[7]. The anticoagulant function of heparin apparently is not involved in this reaction, because the non-anticoagulant fraction of heparin is capable of reversing the anti-angiogenesis activity of protamine. Although coagulation studies remained normal in animals treated with protamine, it is possible that protamine may modify fibrin deposition in the neovascular bed induced by tumours[19,20]. Furthermore, the disparity between the effect of protamine on lung metastases and subcutaneous tumours seems to be due to differences in the tissue uptake of protamine (unpublished data).

These experiments do not prove unequivocally that the angiogenesis-inhibitory activity of protamine is responsible for its anti-tumour effect. However, indirect evidence in support of this conclusion is based on four observations: (1) when protamine is removed from the cornea, capillary growth resumes *before* tumour regrowth begins; (2) histological sections from protamine-treated animals show many microscopic lung metastases that remain avascular, in contrast to saline-treated animals where vascularized tumours predominate; (3) tumour regression in mice does not seem to be immunologically mediated, because lymphocytic infiltration is minimal or absent, and when animals taken off protamine after complete tumour regression are re-inoculated with melanoma cells, tumours grow as before; (4) protamine is neither cytostatic nor cytotoxic to tumour cells *per se* at the doses used.

Four lines of evidence argue that protamine does not act directly on tumour cells: (1) tumour cells in culture grow well at protamine concentrations greater than the peak concentrations of protamine in tumour; (2) the early growth of the protamine-susceptible strain of melanoma (before neovascularization) in protamine-treated mice is not significantly different from that of melanomas in saline-treated mice; (3) the ^{3}H-thymidine-labelling index of bone marrow and viable tumour cells does not differ significantly between protamine- and saline-treated mice; (4) melanoma cells in ascites proliferate during protamine treatment even while the solid primary tumour regresses.

The effectiveness of protamine in inhibiting the growth of lung metastases and its relative ineffectiveness against many subcutaneous tumours might be partly explained on the basis of differential protamine uptake by the two tissues. However, there is no explanation for the finding that strains of tumour implanted s.c. should have different susceptibility to protamine. Some tumours have been found to contain two subpopulations of neoplastic cells—those that can stimulate angiogenesis and those that cannot[21]—and it is possible that the ratio of these two subpopulations dictates the intensity of angiogenesis for a given tumour mass. A second possibility is that different tumour strains might secrete different amounts of glycosaminoglycans that would bind protamine in competition with heparin. A third speculation is that in protamine-resistant tumours there may be more mature vessels. Little is known about the extent to which new capillaries within a tumour convert to mature capillaries and venules (that is, with intact basement membrane, and accumulated pericytes or smooth muscle). Finally, there may be important differences in mast cell content of tumours that would account for different responses to protamine.

Angiogenesis inhibition is a newly found property of protamine. However, this is not the first demonstration of the antitumour activity of protamine, as antitumour activity of protamine or mixtures containing it has been previously found in mice and humans[19,22–26].

It is possible that angiogenesis inhibitors will eventually be used clinically in those diseases common to ophthalmology, dermatology and rheumatology where neovascularization is the dominant pathology, and in certain neoplasms that are the most highly angiogenic, such as brain tumours. It is, however, much too early to know whether protamine or any other inhibitor of angiogenesis, conceivably including some of endogenous origins[27–31], will be of any clinical value. Protamine is toxic

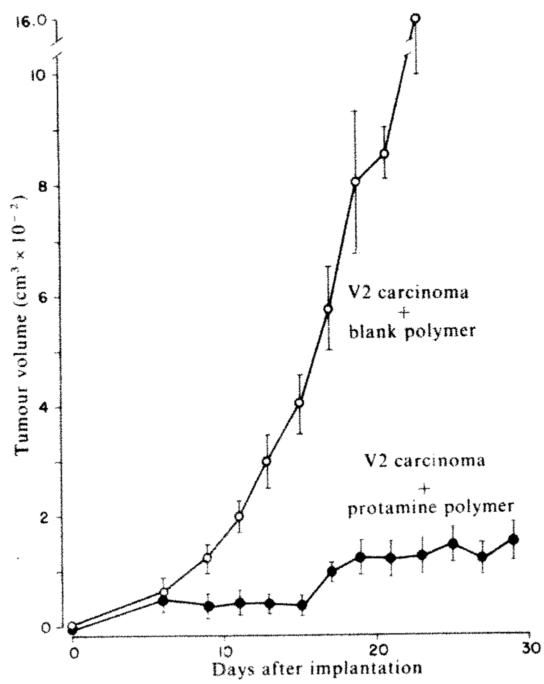


Fig. 5 Inhibition of tumour growth by locally applied protamine when the tumour is implanted directly into the vascular bed. Eight rabbits received a subcutaneous implant of V2 carcinoma in each ear by a method developed by Bronsther *et al.*[15]. One ear received two polymer pellets impregnated with protamine as described in Fig. 3 legend. In the opposite ear, two empty polymer pellets were implanted with the tumour. In both ears, tumour growth formed a hemisphere, the dimensions of which were measured with a micrometer caliper. Tumour growth in the protamine-treated pocket (●) was markedly suppressed compared with tumour in the control ear (○). Although capillary growth rate could not be quantified as in the cornea, the extent of vascularization could be observed by transillumination of the ears. The untreated tumours became vascularized on about days 6–7, and rapid growth ensued. In contrast, the protamine-treated tumours generally did not become vascularized, and their growth was slow or undetectable.

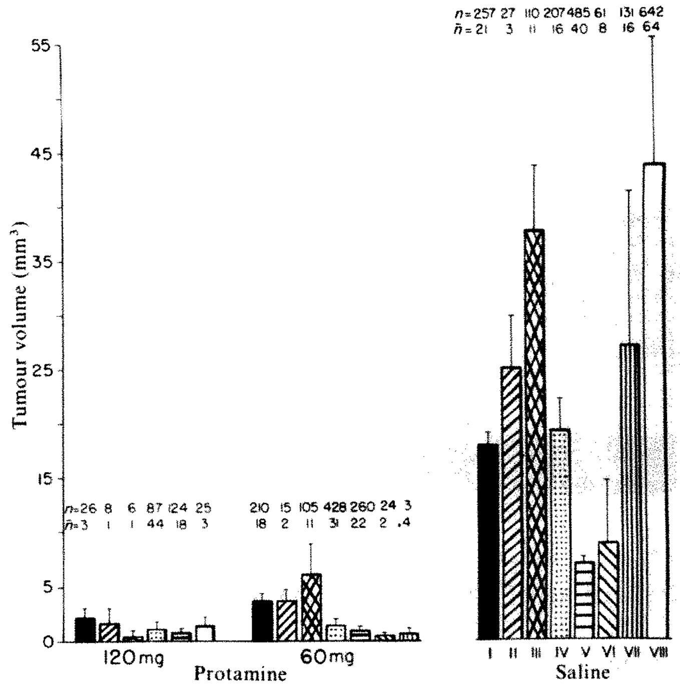


Fig. 6 Effect of protamine of pulmonary metastases in mice and rats, ■, Expt I: 36 C57BL/6 male 42-day-old mice received 1.2×10^6 Lewis lung carcinoma cells i.v., and either protamine, 120 or 60 mg per kg, s.c., every 12 h or saline. All mice were killed on day 15. Mice surviving to that day were: 120 mg, 9/12 (3 mice died free of lung metastases on day 2 presumably from a too rapid step-up to the full dose); 60 mg, 12/12; saline, 12/12. The standard error of the differences in means between all groups was significant at $P<0.001$. In all experiments n = total number of metastases represented by the mean, and $\bar{n}$ = average number of metastases per animal. In this and all subsequent experiments, the mice were weighed twice weekly and the dose of protamine revised upward by increasing the volume according to the average weight of each treatment group except for expts IV and VII where doses were individually calculated. The lungs were fixed in 10% formalin and the lobes separated. The volume and number of metastases were determined with a ×12 stereoscope (eyepiece micrometer 1 unit = 0.064 mm). Two diameters were measured in mm and the volume calculated using the formula[33], vol = $(d_1\ d_2\ d_3\ \Pi)/6$, where the width of the metastasis is used twice as d_1 and d_2 and the length as d_3. The measurements were made by two readers on different days who did not know how each specimen was treated. All mice were housed individually. A stock solution of protamine sulphate (Calbiochem) of 10 mg ml^{-1} in saline was warmed to 37 °C and filtered (0.45 μm Millipore). The amount of protamine lost in the filter was always less than 4% of the original concentration as measured spectrophotometrically at 230 nm. The protamine solution was stored at 4 °C in a sterile 50 ml polyethylene centrifuge tube and warmed to 37 °C before use. Treatment began 24 h after tumour inoculation, and only one injection was given on day 1. The twice daily injections began on day 2. Injections were made with a 30-gauge needle in a shaved area over the iliac crest that had been cleaned with a small amount of Betadine solution. Precautions: if one or two doses of protamine are omitted, resumption of the full dose may be highly toxic or lethal. Stepwise resumption to the full dose is necessary. Male mice comfortably tolerated 60 mg per kg, s.c., 12 hourly for 1 yr without signs of toxicity or without showing differences from saline-injected animals. However, at 120 mg per kg 12-hourly, toxicity appeared after 2–3 weeks in 50% of mice. Tumour-bearing mice show less toxicity for a given dose of protamine than do mice without tumours. If a skin ulcer appears, protamine may leak, resulting in non-therapeutic levels of protamine. If a s.c. injection is given too quickly or forcibly, some protamine will enter the venous system and (with doses > 60 mg per kg) cause sudden toxicity or death. The maximum tolerated single i.v. dose for a mouse is ~6 mg per kg. ▨, Expt II: 56 mice received 6×10^5 Lewis lung carcinoma cells i.v., were placed in three treatment groups and killed on day 17. Surviving mice were: 120 mg, 9/25; 60 mg, 10/16; saline, 10/15. Lung metastases were measured in formalin-fixed lungs as described above. In the protamine-treated mice that died before the end of the experiment, the mean volumes of the pulmonary metastases were less than the mean volumes of the metastases in the killed mice. The s.e. of the difference in means of the 60 mg versus saline groups was significant at $P<0.001$ and that of the 60 mg versus 120 mg groups was significant at $P<0.05$. ▩, Expt III: this experiment was designed to determine the maximum size that could be reached by pulmonary metastases before 50% of the saline-treated host animals had died. Thirty mice received 5.8×10^6 Lewis lung carcinoma cells i.v. and were separated into three groups of 10 mice each. One mouse chosen at random from each protamine group was killed whenever a saline-treated mouse died of tumour, until day 21 when 50% of saline-treated mice had died, and then all the remaining mice were killed. ▒, Expt IV: this experiment was designed to determine the effect of protamine on pulmonary metastases shedding from a primary tumour. Seventy-one mice received a s.c. implant (1 mm³) Lewis lung carcinoma; 25 received protamine, 120 mg per kg every 12 h, 25 received 60 mg and 21 received saline. By day 21 the mean volume of the s.c. primary tumour in the saline-treated mice was 5.4 ± 0.2 cm³; it was 2.4 ± 0.2 cm³ and 2.9 ± 0.6 cm³, respectively, 120 mg and 60 mg groups. All surviving mice were killed on day 24. The survivors were 2/25 in the 120 mg group, 11/25 in the 60 mg group and 7/21 in the saline group. In the protamine-treated mice that died before the end of the experiment, the mean volumes of the pulmonary metastases were less than the mean volumes of the metastases in the killed mice. The s.e. of the difference in means of the 60 mg group versus saline was significant at $P<0.001$. There was no difference between the 60 mg and 120 mg groups. ▤, Expt V: in this repeat of expt IV, 45 mice received s.c. implants of Lewis lung carcinoma and were divided into three groups of 15 mice each. On day 22 the mean volume of the s.c. primary tumour in the saline-treated mice was 6.1 ± 0.3 cm³; it was 3.2 ± 0.2 cm³ and 2.9 ± 0.5 cm³ in the 60 mg and 120 mg groups, respectively. All surviving mice were killed on day 26. The survivors were 7/15, 12/15 and 12/15 for the 120 mg, 60 mg and saline groups, respectively. In the protamine-treated mice that died before the end of the experiment, the mean volumes of the pulmonary metastases were less than the mean volumes of the metastases in the killed mice. The s.e. of the difference in means of the 60 mg group versus saline was significant at $P<0.001$. There was no significant difference between the 60 mg and the 120 mg groups. ▧, Expt VI: this experiment was designed to determine the effect of protamine on pulmonary metastases shedding from a primary tumour which was unresponsive to protamine in the subcutaneous position. Twenty mice received a s.c. inoculation of 10^7 B16 melanoma cells (subcutaneous strain); 12 mice received protamine, 60 mg per kg, s.c., 12 hourly, and 8 received saline. Lung metastases were measured whenever a mouse died. By day 50 three protamine mice and two saline mice were alive and these were killed to end the experiment,. The mean volume of the s.c. tumour at the last measurement before death was 14.4 ± 2.0 cm³ for saline-treated mice and 11.2 ± 1.5 cm³ for protamine-treated mice. The bar graph represents the metastases from all mice. The s.e. of the difference in means between the two groups was significant at $P<0.001$. ▥, Expt VII: a parallel experiment to expt VI except that the primary tumour was responsive to protamine in the subcutaneous position. Sixteen mice received 10^7 B16 melanoma cells (tissure culture strain); 8 mice received protamine, 60 mg per kg 12-hourly, and 8 received saline. Lung metastases were measured whenever a mouse died. By day 50 three mice were alive in each group and they were killed to end the experiment. The mean volume of the s.c. tumour at the last measurement before death was 13.5 ± 1.0 cm³ for the saline-treated mice and 0.3 ± 0.2 cm³ for the protamine group. The protamine-treated tumours underwent either complete or partial regression, but in this experiment, the protamine dose was not lowered during the regression in order to maintain the dose constant for pulmonary metastases. The s.e. of the difference in means between the two groups was significant at $P<0.001$. □, Expt VIII: a test of protamine against pulmonary metastases in a different species. Nineteen CD rats (age 21 days) received 5.1×10^6 Walker carcinoma cells i.v.; nine rats received protamine, 120 mg per kg, s.c., every 12 h and 10 received saline. On day 9, 8 of the 10 saline rats died (due to massive pulmonary metastases) and therefore all remaining rats were killed and lung metastases were measured as before except that Bouin's solution was used for fixation. The bar graph represents the mean volume of metastases from all rats. The s.e. of the difference in means between the two groups was significant at $P<0.001$.

when administered systemically above doses of 60 mg per kg in mice and 120 mg per kg in rats, and this toxicity is unrelated to its angiogenesis-inhibitory activity. This toxicity may severely limit or prevent the clinical use of systemically administered protamine as an antitumour agent. However, in the experimental situations where the appropriate tissue concentration can be achieved, protamine inhibits angiogenesis with the resulting effect on tumours. Furthermore, if the mechanism of the anti-angiogenesis activity of protamine can be determined, this may lead to the discovery of other angiogenesis inhibitors.

We thank Dr Christian Haudenschild for reviewing histological sections; Dr Robert Rosenberg for the non-anticoagulant fractions of heparin; Drs Bruce Zetter and Anthony Bellvé for helpful discussions; Dr Dean Hsieh for the polymer pellets; Ms Martha Potts, Mr Ronald Pelletier, Mr Paul Wesley, Ms Nancy Healy, and Mr Henoy Sims for technical help; Mr Janis Cirulis for art work; Mr Carl Cobb and Ms Jane Dittrich for editorial assistance; Mrs Pauline Breen for typing; Ms Margo Cantor for help in measuring Lewis lung metastases; and Drs Robert Handin for platelet factor 4 and Gerald Gleich for the major basic protein. The B16 melanomas, Lewis lung carcinoma and Walker 256 carcinoma were obtained by courtesy of Dr Randall Johnson and Mr Isidore Wodinsky of Arthur D. Little Co. This work was supported by a grant provided to Harvard University by Monsanto Company and by USPHS grant RO1-CA14019 from the National Cancer Institute.

Note added in proof: Female C57BL/6 mice tolerate higher doses of protamine (up to 120 mg) without toxicity.

Received 13 January; accepted 23 March 1982.

1. Folkman, J. & Cotran, R. S. *Int. Rev. exp. Path.* **16**, 207–248 (1976).
2. Gospodarowicz, D. & Thakaral, K. K. *Proc. natn. Acad. Sci. U.S.A.* **75**, 847–851 (1978).
3. Greenberg, G. R. & Hunt, T. K. *J. cell. Physiol.* **97**, 353–360 (1978).
4. Gimbrone, M. A. Jr, Leapman, S. B., Cotran, R. S. & Folkman, J. *J. exp. Med* **136**, 261–276 (1972).
5. Folkman, J. *Ann. Surg.* **175**, 409–416 (1972).
6. Kessler, D. A., Langer, R. S., Pless, N. A. & Folkman, J. *Int. J. Cancer* **18**, 703–709 (1976).
7. Ausprunk, D. H. & Folkman, J. *Microvasc. Res.* **14**, 53–65 (1977).
8. Folkman, J., Haudenschild, C. C. & Zetter, B. R. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5217–5221 (1979).
9. Zetter, B. R. *Nature* **285**, 41–43 (1980).
10. Azizkhan, R. G., Azizkhan, J. C., Zetter, B. R. & Folkman, J. *J. exp. Med.* **152**, 931–944 (1980).
11. Rosenberg, R. D. & Lam, L. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1218–1222 (1979).
12. Ausprunk, D., Knighton, D. & Folkman, J. *Devl Biol.* **38**, 237–249 (1974).
13. Langer, R. & Folkman, J. *Nature* **263**, 797–800 (1976).
14. Langer, R., Brem, H., Falterman, K., Klein, M. & Folkman, J. *Science* **193**, 70–72 (1976).
15. Bronsther, O., Hsieh, D., Gross, J. & Folkman, J. *Surg. Forum* **32**, 426–428 (1981).
16. Polverini, P. J., Cotran, R. S., Gimbrone, M. A. Jr & Unanue, E. R. *Nature* **269**, 804–806 (1977).
17. Sidky, Y. A. & Auerbach, R. *Science* **18**, 1237–1238 (1976).
18. Fidler, I. J., Gersten, D. M. & Budmen, M. B. *Cancer Res.* **36**, 3160–3165 (1976).
19. Muggleton, P. W., MacLaren, J. G. & Kyke, W. S. C. *Lancet* **i**, 409–10 (1964).
20. Dvorak, H. F., Dvorak, A. M., Manseau, E. J., Wiberg, L. & Churchill, W. H. *J. natn. Cancer Inst.* **62**, 1459–1472 (1979).
21. Azizkhan, R. G. *et al. Surg. Forum* **32**, 424–426 (1981).
22. O'Meara, R. A. Q. & O'Halloran, M. J. *Lancet* **ii**, 613–614 (1963).
23. Hughes, L. E. *Lancet* **i**, 408–409 (1964).
24. Lutton, A. *Lancet* **i**, 768 (1964).
25. Pasternak, N. A., Sheinderovitch, B. A., Kolesova, I. A. & Ermoljeva, I. V. *Antibiotika* **16**, 816–819 (1971).
26. Csaba, G., Acs, Th., Horvath, C. & Kapa, E. *Br. J. Cancer* **14**, 367–375 (1960).
27. Sorgente, N., Kuettner, K. E., Soble, L. W. & Eisenstein, R. *Lab. Invest.* **32**, 217–222 (1975).
28. Brem, H. & Folkman, J. *J. exp. Med* **141**, 427–439 (1975).
29. Langer, R., Conn, H., Vacanti, J., Haudenschild, C. & Folkman, J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4331–4335 (1980).
30. Goren, S. B., Eisenstein, R. & Choromokos, E. *Am. J. Ophthal.* **84**, 305–309 (1977).
31. Patz, A. *et al. Ophthalmology* **85**, 626–637 (1978).
32. Auerbach, R., Kubai, L., Knighton, D. & Folkman, J. *Devl Biol.* **41**, 391–394 (1974).
33. Auerbach, R., Morrissey, L. W. & Sidky, Y. A. *Cancer Res.* **38**, 1739–1744 (1978).

LETTERS TO NATURE

Observation of additional low-degree 5-min modes of solar oscillation

Philip H. Scherrer*, John M. Wilcox*, Jørgen Christensen-Dalsgaard† & Douglas Gough‡

* Institute for Plasma Research, Stanford University, Stanford, California 94305, USA
† National Center for Atmospheric Research, Boulder, Colorado 80307, USA
‡ Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK

By measuring the difference between the shifts in the Fe 5,124 spectrum line from light integrated from a central circular portion of the solar disk and from an annular portion exterior to it, we have detected high-order solar oscillations with degrees $l = 3$, 4 and 5. The frequencies of the octupole modes agree well with the values obtained from whole-disk measurements at the South Pole[1]. A least-squares fit of the observed frequencies to values interpolated between and extrapolated from the predictions of a sequence of solar models with different chemical compositions selects two models. One of these is almost identical to that obtained by a previous fit[15] of modes with $l \leq 2$, and has a helium abundance somewhat greater than 25% by mass.

Most previous observations of low-degree 5-min solar oscillations have been made with light integrated from the entire solar disk. Claverie *et al.*[2–5] and Grec *et al.*[1,6,7] have measured spectrum line shifts, and Deubner[8] and Woodard and Hudson[9] have measured radiant intensity. Such measurements are most sensitive to modes of degree $l = 0$, 1 and 2; the long continuous Doppler observations made at the South Pole have also isolated octupole modes[1].

Our observations are based on a method used by Severny *et al.*[10]. A Babcock solar magnetograph[11] at the Stanford Solar Observatory, which is normally used for measuring Zeeman splitting, was converted into an instrument that measures Doppler shifts with equivalent sensitivity. For this purpose one selects a magnetically insensitive Fraunhofer line (with $g = 0$). Light from a central circular area of the solar disk, with radius $0.5\ R_{\odot}$, is filtered with a right-hand circular polarizer, and light from the annulus between $0.55\ R_{\odot}$ and $0.80\ R_{\odot}$ is filtered with a left-hand polarizer. The wavelength difference between the right and left polarized spectrum lines is measured by the magnetograph. It is interpreted, using the Doppler formula, as the difference between the averaged line-of-sight velocities from the circular and annular portions of the solar disk[12].

Because our apparatus convolves the Doppler signal with a function whose length scale is less than the radius $R_{\odot}$ of the solar disk, the degrees of the modes of oscillation to which it is most sensitive are higher than for whole-disk observations. Our signal should be dominated by modes with $l = 3$, 4 and 5, though contributions from modes with $l = 2$, 6, 9 and 10 should also be detectable[13]. Here we concentrate on the dominant modes.

We obtained good-quality data on 15 days from 17 June to 14 July 1981. The observations were made with 0.1 s time resolution, and were subsequently averaged into 15 s bins. A slow drift, typically of magnitude $1\ \mathrm{ms^{-1}\ h^{-1}}$, was then removed with a high-pass filter. This left data sets over intervals of average duration 9.6 h. Harmonic-amplitude spectra were computed from each data set using a simple least-squares method for finding Fourier coefficients in the frequency range 2.0–4.5 mHz in steps of 1.0 μHz.

Figure 1 shows the average of the 15 spectra. In common with previous observations, power is found to be concentrated between 2.5 and 4.0 mHz, and seems to be generated by a series of oscillators with uniformly spaced frequencies. But whereas whole-disk observations with comparable resolution[2–4,8] display two peaks per interval of ~136 μHz (corresponding to a mode with $l = 1$ and an unresolved pair with $l = 0$ and $l = 2$), here, as expected, there are three; we infer that they correspond to modes with $l = 3$, 4 and 5.

By selecting every third peak (taking due account of peaks that are obviously missing), one might hope to collect sets of modes of like degree. There are no obvious distinguishing differences between the amplitudes and frequencies of the three sets so obtained, as one would expect theoretically. Therefore we cannot identify the modes using our data alone. It is possible, however, to identify the modes in the data obtained at the South Pole, by comparing the uneven frequency distribution and variable peak heights with theoretical expectations[1,6,7,13–16]. The frequencies of the modes inferred to have $l = 3$ are indicated by vertical lines in Fig. 1. They can be seen to correspond with some of the peaks of our spectrum.

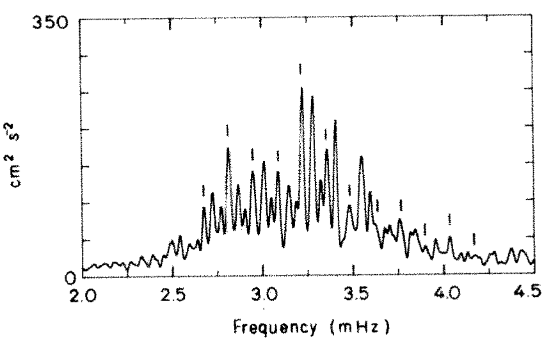


Fig. 1 Average of 15 squared velocity amplitude spectra of data collected in June and July 1981. The vertical lines mark the frequencies of octupole modes measured by Fossat *et al.*[1] at the South Pole.

0028-0836/82/210312—02$01.00

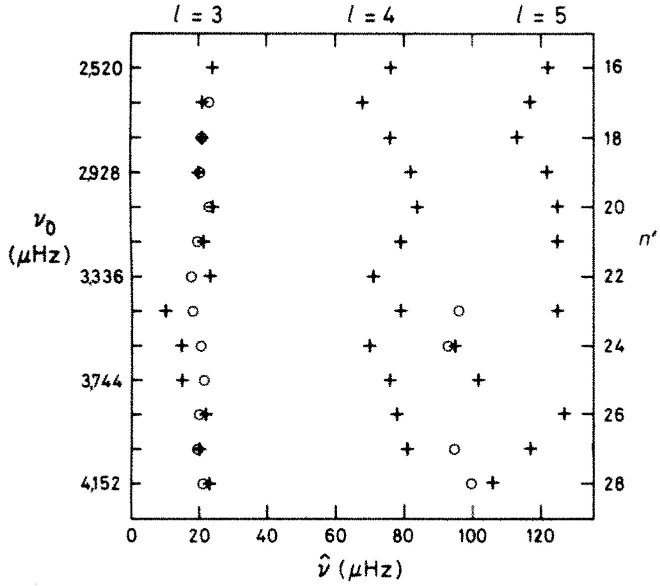


Fig. 2 Echelle diagram of the frequencies between 2.5 and 4.3 mHz of low-degree modes, constructed with $\nu_1 = 72\ \mu$Hz and $\Delta\nu = 136.0\ \mu$Hz. The frequencies ν of the modes represented in the diagram are given by $\nu = \nu_0 + \hat{\nu}$. +, Frequencies of the peaks in Fig. 1. ○, Octupole modes (with $\hat{\nu}$ near 20 μHz) and some dipole modes (with $\hat{\nu}$ near 100 μHz) observed from the South Pole by Fossat *et al.*[1]. Thus by comparing our results with the South Pole data and with theory we identify the degrees of the modes responsible for the three principal columns; these are indicated at the top of the figure. We presume the frequencies 3,703, 3,846 and 4,258 μHz correspond to dipole modes. The ordinate scale n' on the right is the order n or modes with $l \geq 3$, and is $n-1$ when $l = 2$. This identification was obtained by the least-squares comparison with theory described in the text. Fitting regression lines $\alpha + \beta(n - n_0)$, with $n_0 = 22 - \frac{1}{2}l$, to the three principal columns of Stanford frequencies ν yields $(\alpha, \beta) = (3{,}152 \pm 1, 135.7 \pm 0.3)$, $(3{,}140 \pm 1, 136.2 \pm 0.4)$ and $(3{,}117 \pm 1, 136.5 \pm 0.3)$ for $l = 3$, 4 and 5 respectively.

To facilitate identification of the modes responsible for our data, we present, following Grec[17], an echelle diagram (Fig. 2) of the frequencies of the peak maxima in Fig. 1. Each frequency ν is reduced to a frequency $\nu_0 + \hat{\nu}$ where ν_0 is an integral multiple of $\Delta\nu$ plus some constant ν_1, and $\hat{\nu}$ is in the range $(0, \Delta\nu)$. Figure 2 shows a plot of $\nu_0 \equiv \nu - \hat{\nu}$ against $\hat{\nu}$, the spacing $\Delta\nu$ having been chosen to give roughly the same values of $\hat{\nu}$ to modes of like degree. Thus $\Delta\nu$ is representative of the mean frequency differences between modes of adjacent order and like degree. More accurate estimates of these differences are given in the legend.

Included in Fig. 2 are the octupole modes identified in the South Pole data. Because these agree well with one of our columns of frequencies, we deduce that this column also results from octupole modes. The degrees of the new modes producing the other two principal columns are then inferred by comparison with theory[13]. Further comparisons with theory and with other observations will be reported elsewhere[18].

The addition of these new modes to the helioseismological data should impose tighter theoretical constraints on solar models. Here we simply compare the frequencies we have measured with the frequencies ν_λ interpolated between and extrapolated beyond the frequencies ν_A, ν_B, ν_C of the three solar models A, B, C of Christensen-Dalsgaard *et al.*[19] according to the formula

$$\nu_\lambda = \begin{cases} \lambda\nu_A + (1-\lambda)\nu_B & \lambda > 0 \\ (1+\lambda)\nu_B - \lambda\nu_C & \lambda < 0 \end{cases} \quad (1)$$

Model A is a standard solar model with initial helium and heavy-element abundances $(Y, Z) = (0.25, 0.02)$; models B and C have initial abundances (0.19, 0.004) and (0.16, 0.001) respectively, and have suffered contamination by heavy elements in and above the convection zone at a rate chosen to yield a present surface abundance: $Z = 0.02$. The least-squares fit of ν_λ with our frequencies ν represented in Fig. 2 yields $\lambda = -0.29$ and $\lambda = 1.26$ as the two best solutions. The r.m.s. difference between ν_λ and the observations is ~8 μHz for both solutions. The helium-rich solution agrees with that of similar analyses of whole-disk data[15] and low-order modes apparently detected as oscillations in the limb-darkening function[20]. Therefore we select this as our best estimate. We are then able to infer the orders n of the modes; these are indicated in Fig. 2. Our extrapolated solution has a relatively deep convection zone, and is thus consistent with analyses of 5-min oscillations of high degree[21,22].

We thank E. Fossat, G. Grec and M. Pomerantz for communicating observations before their publication and for suggesting the format of Fig. 2. This work was supported in part by the Office of Naval Research under contract N00014-76-C-0207, by NASA under grant NGR 05-020-559 and contract NAS5-24420, by the Solar-Terrestrial Research Program of the Division of Atmospheric Sciences of the NSF under grant ATM77-20580, by the Max C. Fleischmann Foundation, by the National Center for Atmospheriic Research, which is sponsored by the NSF, and by the SERC.

Received 11 March; accepted 13 April 1982.

1. Fossat, E., Grec, G. & Pomerantz, M. *Proc. IAU Colloq.* No. 66 (in the press).
2. Claverie, A., Isaak, G. R., McLeod, C. P., van der Raay, H. B. & Roca Cortes, T. *Nature* **282**, 591–594 (1979).
3. Claverie, A., Isaak, G. R., McLeod, C. P., van der Raay, H. B. & Roca Cortes, T. *Astr. Astrophys.* **91**, L9–L10 (1980).
4. Claverie, A., Isaak, G. R., McLeod, C. P., van der Raay, H. B. & Roca Cortes, T. *Solar Phys.* **74**, 51–57 (1981).
5. Claverie, A., Isaak, G. R., McLeod, C. P., van der Raay, H. B. & Roca Cortes, T. *Nature* **293**, 443–445 (1981).
6. Grec, G., Fossat, E. & Pomerantz, M. *Nature* **288**, 541–544 (1980).
7. Fossat, E., Grec, G. & Pomerantz, M. *Solar Phys.* **74**, 59–63 (1981).
8. Deubner, F.-L. *Nature* **290**, 682–683 (1981).
9. Woodard, M. & Hudson, H. *Proc. IAU Colloq.* No. 66 (in the press).
10. Severny, A. B., Kotov, V. A. & Tsap, T. T. *Nature* **259**, 87–89 (1976).
11. Babcock, H. W. *Astrophys. J.* **118**, 387–396 (1953).
12. Dittmer, P. H. *Astrophys. J.* **224**, 265–275 (1978).
13. Christensen-Dalsgaard, J. & Gough, D. O. *Mon. Not. R. astr. Soc.* **198**, 141 (1982).
14. Christensen-Dalsgaard, J. & Gough, D. O. *Nature* **288**, 544–547 (1980).
15. Christensen-Dalsgaard, J. & Gough, D. O. *Astr. Astrophys.* **104**, 173–176 (1981).
16. Christensen-Dalsgaard, J. *Proc. 5th Eur. Meet. in Astronomy* (ed. P. Ledoux) A.1.1. (Institut d'Astrophysique, Liège, 1980).
17. Grec, G. thesis, Univ. Nice (1981).
18. Scherrer, P. H., Wilcox, J. M., Christensen-Dalsgaard, J. & Gough, D. O. *Proc. IAU Colloq.* No. 66 (in the press).
19. Christensen-Dalsgaard, J., Gough, D. O. & Morgan, J. G. *Ast. Astrophys.* **73**, 121–128, **79**, 260 (1979).
20. Gough, D. O. *Nature* (submitted).
21. Lubow, S. H., Rhodes, E. J. Jr & Ulrich, R. K. in *Nonradial and Nonlinear Stellar Pulsation* (eds Hill, H. A. & Dziembowski, W. A.) 300–306 (Springer, Heidelberg, 1980).
22. Berthomieu, G., *et al.* in *Nonradial and Nonlinear Stellar Pulsation* (eds Hill H. A. & Dziembowski, W. A.) 307–312 (Springer, Heidelberg, 1980).

Jovimagnetic secular variation

J. E. P. Connerney & M. H. Acuña

NASA/Goddard Space Flight Center, Laboratory for Extraterrestrial Physics, Planetary Magnetospheres Branch, Greenbelt, Maryland 20771, USA

Long-term variations of a planetary magnetic field are one of the few observables available for the study of planetary interiors and dynamo theory. Although variations of the geomagnetic field have been accessible to direct measurement for centuries, knowledge of the secular variations of other planetary dynamos is limited. We report here new limits on jovimagnetic secular variations found by comparison of a jovian internal field model[1] obtained from the Voyager 1 magnetic field observations at epoch 1979.2 with the epoch 1974.9 Pioneer 11 O_4 model[2]. No significant secular variation of either the magnitude or position of the jovidipole was found for the years 1974.9 to 1979.2, although a small Earth-like variation cannot be ruled out.

A gradual decrease in the Earth's dipole moment at a rate of ~5% per 100 yr has been established from observations of the geomagnetic field over the past 150 yr (ref. 3). Less well established but still generally accepted is the slow westward

drift of certain features of the geomagnetic field at a rate of ~0.1° or 0.2° long. yr^{-1}. From the palaeomagnetic records it is well known that the Earth's dipole field has a rich history of wanderings and polarity reversals. The irregular and unpredictable switching of the geodipole occurs with varying frequency, a rate of several reversals per Myr being typical of the past 200 Myr.

Geomagnetic secular variations are an important source of information about the Earth's deep interior and the dynamo presumed responsible for the geomagnetic field[4-6]. Within the context of modern dynamo theory, secular variations can provide constraints on material properties deep within the planet's otherwise inaccessible interior. Comparative studies of planetary dynamos may prove invaluable in unravelling the complexities of the self-sustaining dynamo. Thus the history of Jupiter's planetary magnetic field is of great interest, especially because it represents the first real opportunity to observe secular variations of a planetary dynamo other than the Earth's.

Indirect observations of the jovian magnetic field began with the discovery of non-thermal radio emissions by Burke and Franklin[7] in 1955 and are still in progress[8]. Continued observations of jovian radio emissions have precisely established the rotation rate of the planet and have suggested the occurrence of secular variations of the jovimagnetic field. Berge[9] has suggested that the 30% decrease in Jupiter's integrated flux density occurring from 1961 to 1973 may be a result of a decrease in the jovian magnetic dipole moment. An upper limit on the variation of magnetic field strength between epochs 1967 and 1970 of ~5% yr^{-1} was obtained from observations of the circular polarization of decimetric radiation[10]. From observations of the maximum frequency of jovian decametric radiation, J. K. Alexander (personal communications) concludes that the magnetic field magnitude at high northern latitudes has changed by <0.1% yr^{-1}. Observations relating to the inclination of the jovidipole with respect to the planet's rotation axis have been interpreted as evidence of a secular decrease of the inclination angle by $0.07 \pm 0.05°$ yr^{-1} (ref. 10) but have not been confirmed[11].

In situ observations of the jovian planetary magnetic field are limited to the Pioneer 10 and 11 flybys at epochs 1973.9 and 1974.9 and the Voyager 1 and 2 flyby encounters at epochs 1979.2 and 1979.5. Comparison of spherical harmonic models of the jovimagnetic field obtained from the Pioneer observations[12] suggested a decrease in the dipole moment of ~6% during the intervening year. The decrease was, however, attributed to other effects although the possibility of a secular change as large as 6% per year was explicitly not excluded[13].

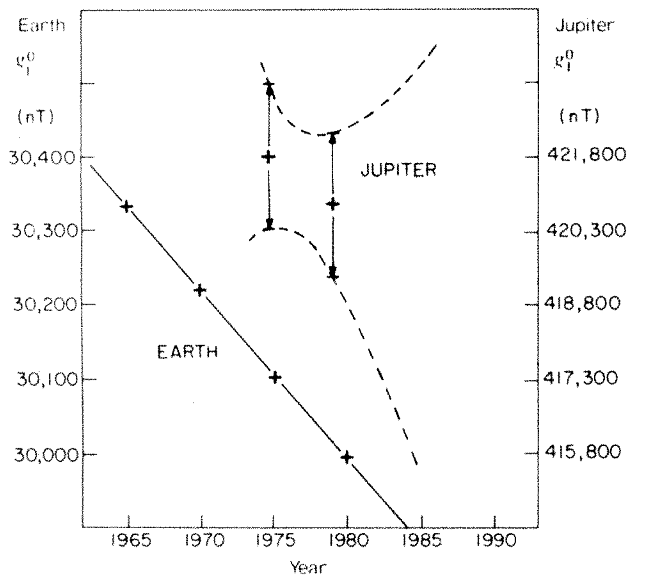


Fig. 1 Secular decrease of the Earth's main dipole g_1^0 term compared with estimates of Jupiter's main dipole term. One vertical division represents ~0.033% of the main dipole term in each case.

Table 1 Jovian magnetic field models

	(1979.2) V1 17 ev*	(ε)	P11 O^4 (GSFC)†	(2σ)‡	(1974.9) P11 15 evs‡	P11 SHA 23§
1. g_1^0	4.208	(0.032)	4.218	(0.030)	4.068	4.092
2. g_1^1	−0.660	(0.004)	−0.664	(0.019)	−0.668	−0.705
3. h_1^1	0.261	(0.005)	0.264	(0.024)	0.243	0.231
4. g_2^0	−0.034	(0.028)	−0.203	(0.023)	−0.093	−0.033
5. g_2^1	−0.759	(0.030)	−0.735	(0.039)	−0.672	−0.699
6. g_2^2	0.483	(0.018)	0.513	(0.048)	0.502	0.537
7. h_2^1	−0.294	(0.058)	−0.469	(0.037)	−0.498	−0.531
8. h_2^2	0.107	(0.018)	0.088	(0.037)	0.119	0.074
9. g_3^0	—	(—)	−0.233	(0.060)	−0.111	−0.113
10. g_3^1	—	(—)	−0.076	(0.083)	−0.316	−0.585
11. g_3^2	0.263	(0.114)	0.168	(0.080)	0.220	0.283
12. g_3^3	−0.069	(0.054)	−0.231	(0.082)	−0.250	0.067
13. h_3^1	—	(—)	−0.580	(0.104)	−0.476	−0.423
14. h_3^2	0.695	(0.108)	0.487	(0.108)	0.380	0.120
15. h_3^3	−0.247	(0.054)	−0.294	(0.068)	−0.228	−0.171

1965 system III, ϕ positive east unless noted. Schmidt normalized spherical harmonic coefficients, Gauss.

* Ref. 1. † Ref. 2 rotated to 1965 system III. ‡ Ref. 16. § Ref. 12 rotated to 1965 system III.

A preliminary jovimagnetic field model based on Voyager 1 observations[14] yielded a similarly reduced dipole moment attributed not to a secular decrease of the jovian field but rather to the presence of a large-scale current system in the jovian magnetosphere. This azimuthally directed 'magnetodisk' current is confined to an equatorial annular disk extending from ~5R_J (jovian radii) to >50R_J (ref. 15) and appears to be a permanent feature of the voluminous jovian magnetosphere. The relative contribution of this ~3×10^8 A current system to the magnetic field observed by Voyager 1 near closest approach is greater than that of previous flybys by virtue of the larger periapsis of the Voyager 1 encounter (4.9 R_J) relative to those of Pioneer 10 and 11 (2.8 and 1.6 R_J).

A more recent estimate of the jovimagnetic field has been obtained from the Voyager 1 observations[1] through the use of new modelling and inversion techniques[16] which facilitate the separation of externally generated fields from those generated within Jupiter's core. The Schmidt normalized spherical harmonic coefficients of this recent Voyager 1 internal field model are listed in Table 1 along with some previous estimates obtained from the Pioneer 11 vector helium[12] and fluxgate magnetometer observations[2]. The parameters which have no corresponding entry in the V1 list are not determinable from the Voyager 1 magnetic field observations alone and are therefore not available at epoch 1979.2. There is general agreement among the Pioneer 11 and V1 models, and the correspondence between the dipole terms of the V1 and O$_4$ models is particularly good. Here we consider only a comparison of the V1 and Pioneer 11 O$_4$ model parameter estimates and the associated estimated parameter uncertainties (in parentheses in Table 1). (Davis and Smith[13] estimate that uncertainties in the dipole terms of the SHA 23 model quoted in Table 1 are as large as 5% of the g_1^0 term, that is, ~0.2 G. They estimate that uncertainties in the quadrupole and octupole terms range from 0.2 to 0.8 G. Since these uncertainties are approximately an order of magnitude greater than those estimated for the O$_4$ and V1 models we do not consider the SHA 23 model.)

In Fig. 1, the time dependence of the magnitude of the g_1^0 coefficient for both the Earth[17] and Jupiter are compared, scaled such that a major division represents ~0.33% of the field magnitude. A linear fit to the jovian g_1^0 estimates shown in Fig. 1 is

$$g_1^0(t) = 4.218\ (\pm 0.015) - 0.0023\ (\pm 0.0050)t \qquad (1)$$

where g_1^0 is given in G, and t is the time difference in yr from

1974.9. No significant secular variation of the magnitude of the g_1^0 term is found. The observations are consistent with a modest secular increase of ~0.06% yr (at 1 s.d. in the rate) or a decrease of as much as ~0.17% yr^{-1}. For comparison, the observed secular decrease of the magnitude of the Earth's g_1^0 term is ~0.075% yr^{-1}. The dashed lines in Fig. 1 correspond to a 1 s.d. error in an estimate of g_1^0 calculated using equation (1), illustrating the deterioration in predictive capability for epochs far removed from the 1977 median epoch.

A comparison of the two remaining dipole terms g_1^1 and h_1^1 leads to a negligible inferred drift of the magnetic dipole axis of 0.025° (±0.22) yr^{-1} in longitude. The estimated uncertainty in the drift rate is comparable with the uncertainties in Jupiter's rotation rate[18]. The inclination of Jupiter's dipole axis has decreased by an insignificant ~0.01° (±0.03) yr^{-1}. These conclusions are not substantially altered if an alternative model[16] derived from the Pioneer 11 fluxgate magnetometer observations is used in place of the O_4 model. The uncertainties quoted here are largely the result of uncertainties in the position of the dipole axis at epoch 1974.9. Some of the V1 quadrupole and octupole parameters (in particular the g_2^0, h_2^1, g_3^2 and h_3^2 terms) depend on the details of the model used to represent the jovian magnetodisk currents[1,15] and therefore we will not discuss higher order terms at present.

The *in situ* magnetic field observations are consistent with no secular variation of the jovimagnetic dipole field from 1974.9 to 1979.2. They are also consistent with a modest Earth-like secular variation in both the magnitude of the main dipole g_1^0 term and the drift of the dipole axis in longitude. Earlier suggestions of relatively large secular variations in the dipole moment and inclination[10] are thus not substantiated. Current theoretical estimates of jovimagnetic secular variations based on free hydromagnetic oscillations of the liquid core[19] range from days to centuries. Assuming the Voyager 1 and O_4 models do not represent a chance agreement of two samples of an aliased time sequence, jovimagnetic secular variations must have a time scale of centuries and not decades. It is also clear that the differences in the Pioneer 10 and 11 internal field models reflects more the influence of unmodelled current systems than a real secular variation of the jovimagnetic field. Estimates of the size of Jupiters' conducting core[20] based on Hide's method[21] and the large secular variations inferred from the Pioneer models need to be critically reassessed.

Finally, the methods used to obtain an estimate of the jovian internal field from the Voyager 1 observations[1] should be equally applicable to flybys planned for the near future with similarly large periapses. A similar determination of Jupiter's internal field from either the International Solar Polar Mission (to ~6 R_J) in 1987 or the Galileo encounter (~5 R_J) scheduled for ~1990 could in principle distinguish between an Earth-like and no secular variation of Jupiter's main field.

Received 18 March; accepted 7 April 1982.

1. Connerney, J. E. P., Acuña, M. H. & Ness, N. F. *J. geophys. Res.* **87** (in the press).
2. Acuña, M. H. & Ness, N. F. *J. geophys. Res.* **81**, 2917–2922 (1976).
3. Rikitake, T. *Electromagnetism and the Earth's Interior* (Elsevier, Amsterdam, 1966).
4. Gubbins, D. *Rev. Geophys. Space Phys.* **12**, 137–154 (1974).
5. Inglis, D. R. *Rev. Modern Phys.* **53**, 481–496 (1981).
6. Braginsky, S. I. *Izv., Earth Physs.* **14**, 659–668 (1978).
7. Burke, B. F. & Franklin, K. L. *J. geophys. Res.* **60**, 213–217 (1955).
8. Carr, T. D., Desch, M. D. & Alexander, J. K. *Physics of the Jovian Magnetosphere* (ed. Dessler, A. J.) (in the press).
9. Berge, G. L. *Astrophys. J.* **191**, 775–784 (1974).
10. Stannard, D. & Conway, R. G. *Icarus* **27**, 447–452 (1976).
11. Neidhofer *et al. Astr. Astrophys.* **61**, 321–328 (1977).
12. Smith, E. J., Davis, L. Jr & Jones, D. E. *Jupiter* (ed. T. Gehrels) (University of Arizona Press, Tucson, 1976).
13. Davis, L. Jr & Smith, E. J. *magnetospheric Particles and Fields* (ed. McCormack, B. M.) 301–310 (Reidel, Dordrecht, 1976).
14. Ness, N. F. *et al. Science* **204**, 982–987 (1979).
15. Connerney, J. E. P., Acuña, M. H. & Ness, N. F. *J. geophys. Res.* **86**, 8370–8384 (1981).
16. Connerney, J. E. P. *J. geophys. Res.* **86**, 7679–7693 (1981).
17. IAGA Division 1 Working Group 1 *EOS* **62**, 1169 (1981).
18. May, J., Carr, T. D. & Desch, M. D. *Icarus* **40**, 87–93 (1979).
19. Hide, R. & Stannard, D. *Jupiter* (ed. Gehrels, T.) 767–787 (University of Arizona Press, Tucson, 1966).
20. Hide, R. & Malin, S. R. C. *Nature* **280**, 42–43 (1979).
21. Hide, R. *Nature* **271**, 640–641 (1978); *Geophys. Astrophys. fluid Dyn.* **12**, 171–176 (1979).

Molecular relaxations in a glass of cholesteric liquid crystal

G. P. Johari*, J. W. Goodby & G. E. Johnson

Bell Laboratories, Murray Hill, New Jersey 07974, USA

Glasses, in which the orientations of molecules have a long-range order, can be obtained by supercooling the mesophases of liquid crystals. Such mesophases may be: (1) nematic, when the centres of their rod-, or lath-like molecules do not lie on a regular lattice, but there is a long-range order in the alignment of their long axes, called the director; (2) cholesteric (or twisted nematic) when the alignment (due to steric asymmetry of the individual molecules) of the nematic type is such that the local director rotates at a steady rate as one moves normal to it; and (3) smectic, when there is also a partial ordering in the positions of the molecules in planes related to the local direction of orientational order. The situation in the nematic and cholesteric phases is clearly akin to magnetic ordering and the glassy states of these phases are useful, as our models of disorder, in understanding the behaviour of structurally isotropic glasses. A study of molecular motions in the glassy cholesteric phase of cholesteryl hydrogen phthalate, given here, shows features which have a striking resemblance to those observed in the (positionally ordered) glass-like state of plastic crystals[1,2] and the (grossly disordered) molecular and polymeric glasses. This suggests that the characteristic features of the molecular motions in glasses must be explained in terms of the non-equivalence of the molecular environment, rather than in terms of the complexity of the molecule, its internal degrees of freedom, or the state of aggregation. This has implications for our concept of the microstructure of a glass.

Cholesteryl hydrogen phthalate melts to an isotropic liquid at 431 K. The liquid readily supercools and undergoes an isotropic→cholesteric transition at 367 K. On further cooling the turbid cholesteric phase becomes increasingly more viscous and undergoes a glass transition at 297 K, as indicated by differential thermal analysis. In the cholesteric phase[3] the alignment of the molecular axes has a long-range order as in a nematic phase, but this order is restricted to within the plane and each subsequent plane has molecular orientations (or local director) twisted by an angle of <1°. Thus the arrangement acquires a helical pattern with the optical axis coincident with the helical axis, which is normal to the planes. Measurements were made on unoriented samples contained in a parallel-plate, three-terminal dielectric cell whose temperature was controlled from 77 to 440 K. The distance between the highly polished surfaces of the electrodes was ~2 mm and the strength of the electric field was 20–25 V cm^{-1}. The cholesteric phase between the electrodes had no preferred orientation; the symmetry of the sample was helical over distances of, say, a few micrometres and spherical over longer distances, because the order induced by the surface of electrodes extends to a distance of a few hundred micrometres, and thus the helical axes for the domains may assume all possible orientations. The capacitance and conductance, or dielectric loss factor, were measured from 0.1 to 100 Hz, using a low-frequency bridge[4], and from 10^2 to 10^5 Hz, using a GR 1615A capacitance bridge. No change in capacitance or conductance occurred with a change in the electric field strength from 2.5 to 25 V cm^{-1}.

The dielectric loss tangent, tan δ, at 1 kHz of the supercooled and glassy cholesteric phase is plotted against temperature in Fig. 1. The two peaks are due to molecular relaxations, in the supercooled liquid at 321 K and in the glass at 155 K whose

* Permanent address: Snow and Ice Division, National Hydrology Research Institute, Ottawa, Canada K1A 0E7.

0028-0836/82/210315—03$01.00

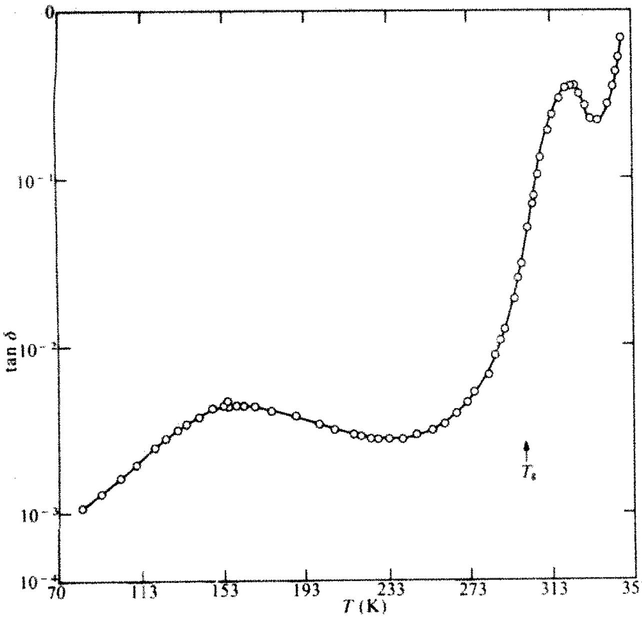


Fig. 1 The dielectric loss factor, tan δ, at 1 kHz of the unoriented samples of cholesteric phase of cholesteryl hydrogen phthalate in the supercooled liquid and glassy states plotted against temperature. T_g is the calorimetric glass transition temperature.

rates correspond to the frequency of 1 kHz. These are referred to as main and secondary relaxations respectively. The presence of a secondary relaxation indicates that, even in the otherwise rigid matrix of a cholesteric glass, configurational states which involve changes in the electric polarization are available to the structure. The lower magnitude of its dielectric loss, or orientation polarization, also suggests that either the number of molecules contributing to the secondary relaxation and/or the dipole moment associated with the molecular reorientations is much lower than that in the main relaxation.

The spectrum of the main relaxation in a frequency plane is asymmetric and is broader at the high-frequency than at the low-frequency side of the peak. The half-width of the peaks is ~2 decades of frequency, which is significantly higher than 1.14 decades characteristic of a single Debye-type relaxation. The spectrum of the secondary relaxation given in Fig. 2 is also broad. Here the half-width of the peaks increases from 4 decades at 190 K to 5 decades at 139 K. Thus both the main and secondary relaxations have a distribution of relaxation times, but the distribution changes significantly with temperature only for the latter. The relaxation rates, which are given by the frequency, f_m, of the peaks in the spectrum, are plotted logarithmically against the reciprocal temperature in Fig. 3. The rate of the main relaxation follows the Vogel–Fulcher–Tamman equation of the type, $f_m = A \exp(-B/(T-T_0))$, where A, B and T_0 are empirical contants. The extrapolated relaxation rate at the calorimetric T_g (= 297 K) is $10^{-4}\ s^{-1}$. Thus the molecular degrees of freedom causing the dielectric relaxation of the cholesteric phase are undoubtedly the ones whose freezing out causes the decrease in the heat capacity at T_g. The rate of the secondary relaxation in Fig. 3 follows the Arrhenius equation, $f_m = f_0 \exp(-E/RT)$. This gives an activation energy, E, of 19.8 kJ mol^{-1} for the molecular processes in the cholesteric glass.

The aforegiven features of both the main and secondary relaxations in the cholesteric liquid crystal are strikingly similar to those observed in amorphous polymers[5,6], rigid molecular glasses[7] and orientationally disordered crystals[2]. This similarity suggests that the potential energy barriers resisting the molecular motions, whose freezing out causes the glass transition, are independent of the internal degrees of freedom and the state of aggregation of molecules, and that the molecular mobility is intrinsic to the nature of the glassy state.

In terms of the models for the structure of a glass, which involve dense random packing[8,9], microcrystals[10], dislocations[11–14], mixed clusters of competing polymorphs[15,16], disclinations[17] and dislocations[18], the secondary relaxation may arise from hindered rotations of molecules in localized regions of loose molecular packing[19] (referred to as 'islands of mobility') bounded by the relatively immobile close-packed structures. The localized regions may consist of the interstitial packing[20] between the amorphous clusters, or polyhedra, if the structure of a glass consists of regions of varying molecular density. In such a case only a fraction of the total number of molecules in a glass may contribute to its secondary relaxation. Although such localized regions may exist due to the freezing-in of the short range structures corresponding to the density fluctuations in the liquid at T_g, it is also believed that the secondary relaxation indicates small angle movement of each molecule[2] in the essentially equivalent molecular environments of a dense random packed structure of a glass. In this case all molecules in a glass contribute to the relaxation but each by a small amount.

Owing to the nature of the packing of the rod- or lath-like molecules fewer regions of relatively loose structure would exist in a glass obtained from a cholesteric liquid than in one obtained from an isotropic liquid. Consequently, the magnitude of tan δ in the secondary relaxation relative to the main would, in general, be lower in the former than in the latter type of glass. But, if the small angle movement of all molecules were responsible for the secondary relaxation, as is expected in the dense random packed structure of a glass, the aforementioned relative magnitude of tan δ would be similar in the two types of glasses. The ratio of maximum tan δ at 1 kHz of the secondary to the main relaxation in Fig. 1. is 1 : 58 in the cholesteric glass, which is much lower than the corresponding ratio of 1 : 10 or less, observed in most molecular[2,7] and polymeric glasses[5]. Therefore, the secondary relaxation is likely to be associated with the regions of relatively loose interstitial packing in the strucure of a glass.

These results are significant for our understanding of the molecular processes in disordered solids, for it seems that the features of molecular motions in them must have their origins in the non-equivalence of the molecular environment, rather than in the complexity of the molecules, their internal degrees of freedom, or the state of their orientational or positional disorder. The results also support a model for the structure of a glass as an assembly of polyhedra connected by loose interstitial packing, where molecular motion is possible even when the polyhedra are fixed in place. Because supercoooled thin

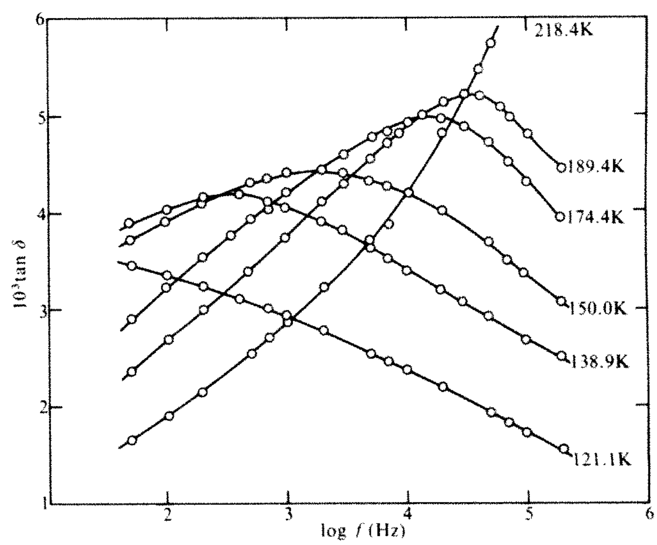


Fig. 2 The dielectric loss factor, tan δ, plotted against frequency for the secondary relaxation in the glassy cholesteric phase at several temperature.

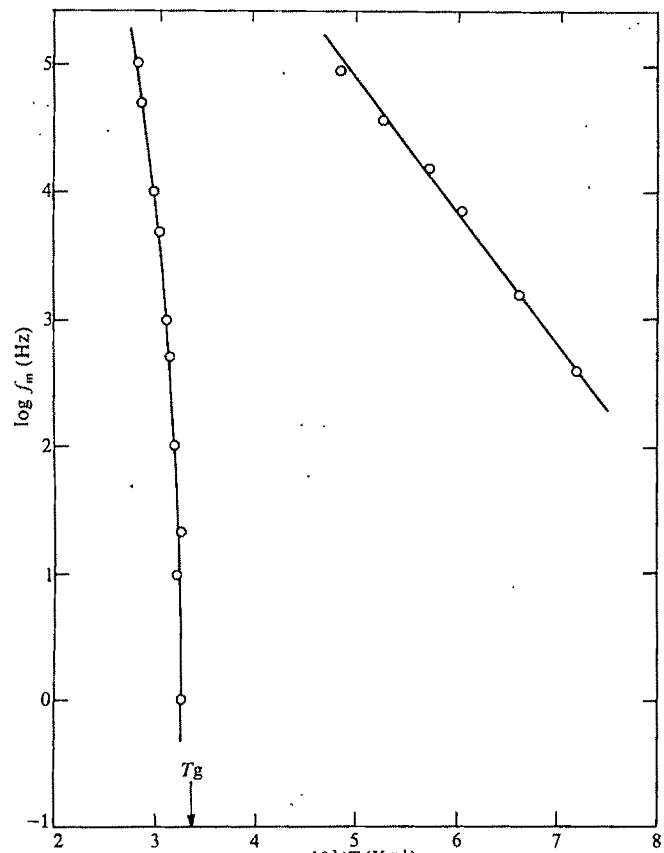


Fig. 3 The frequency of maximum loss, f_m, plotted logarithmically against the reciprocal temperature for main relaxation, and secondary relaxation, in the supercooled and glassy cholesteric phase. T_g is the calorimetric glass transition temperature.

films of cholesteric glass in our experiments are found to be anisotropic, we maintain that the polyhedra in the structure are not microcrystals or crystal nuclei arrested in various stages of their growth, for a random arrangement of such polyhedra would produce a macroscopically isotropic, not anisotropic, structure.

The results have a further implication for our understanding of the glass transition in metal alloys and of the computer simulation of hard sphere glasses, for the configurational restrictions that lead to the formation of a glass in a mesomorphic phase involve mainly the positions (centres of mass) of the molecules and not the orientations or the local director, as no change in the order parameter occurs at T_g. In this respect the glass transition in them is similar to the glass transition in metallic alloys and in the computer simulated experiments on aggregates of hard spheres, where the choice of spheres precludes the possibility of an orientational disorder.

Received 13 January; accepted 1 April 1982.

1. Suga, H. & Seki, S. *J. Noncryst. Solid* **16**, 171–194 (1974).
2. Johari, G. P. *Ann. N.Y. Acad. Sci.* **279**, 117–140 (1976); *Phil. Mag.* **41**, 41–47 (1980).
3. Gray, G. W. in *Liquid Crystals and Plastic Crystals* Vol. I, Ch. 4 (eds Gray, G. W. & Winsor, P. A.) (Harwood, Chichester, 1974).
4. Johari, G. P. & Jones, S. J. *Proc. R. Soc.* **A349**, 467–495 (1976).
5. McCrum, N. G., Read, R. E. & Williams, G. *Anelastic and Dielectric Effects in Polymeric Solids* (Wiley, London, 1967).
6. McCall, D. W. *Nat. Bur. Stand. U.S. Spec. Publ.* **301**, 475–537 (1969).
7. Johari, G. P. & Goldstein, M. *J. Chem. Phys.* **53**, 2372–2388 (1970); **55**, 4245–4252 (1971).
8. Bernal, J. D. *Proc. R. Soc.* **A280**, 299–322 (1964).
9. Finney, J. L. *Nature* **266**, 309–314 (1977).
10. Turnbull, D. & Polk, D. E. *J. Noncryst. Solids* **8–10**, 19–25 (1972).
11. Gilman, J. J. *J. appl. Phys.* **44**, 675–679 (1973).
12. Ashby, M. F. & Logan, J. *Scr. Metallurg.* **7**, 513–521 (1973).
13. Li, J. C. M. in *Frontiers of Science* (eds Murr, L. E. & Stein, C.) 527–562 (Dekker, New York, 1976).
14. Edwards, S. F. & Warner, M. *Phil. Mag.* **40A**, 257–278 (1979).
15. Goodman, C. H. L. *Nature* **257**, 370–372 (1975).
16. Wang, R. & Mertz, M. D. *Nature* **260**, 35–36 (1976).
17. Rivier, N. *Phil. Mag.* 40A, 859–868 (1979).
18. Morris, R. C. *J. Appl. Phys.* **50**, 3250–3257 (1979).
19. Johari, G. P. *J. chem. Phys.* **58**, 1766–1770 (1973).
20. Goldstein, M. *J. Phys. Paris Colloq.* **C2**, 2–3 (1975).

A 2-D model calculation of atmospheric lifetimes for N_2O, CFC-11 and CFC-12

Malcolm K. W. Ko & Nien Dak Sze

Atmospheric and Environmental Research, Inc.,
840 Memorial Drive, Cambridge, Massachusetts 02139, USA

Accurate values for the atmospheric lifetimes of N_2O, $CFCl_3$ (CFC-11) and CF_2Cl_2 (CFC-12) are vital[1–3] for studies of their life cycles and in attempts to forecast their future atmospheric levels. These gases are known to be removed mainly by stratospheric photolysis. The lifetimes set by these removal processes are usually inferred from one-dimensional (1-D) models. Using a two-dimensional (2-D) zonal-mean model together with recent observations, we argue here that the stratospheric removal rates may also depend on certain latitudinal features, including latitudinal variations of the gas concentrations, photolysis rates and their correlations, features that cannot be included in the 1-D models. Our discussion also considers the validation of 1-D model results, particularly regarding predictions of atmospheric lifetimes.

The atmospheric lifetime (τ) of a trace gas in steady-state condition can be defined as

$$\tau = \frac{\langle \text{global burden} \rangle}{\langle \text{global removal} \rangle} = \frac{\langle \int fN \, dv \rangle}{\langle \int LNf dv \rangle} \qquad (1)$$

where N is the air number density, f and L are the steady-state volume mixing ratio and loss frequency of the trace gas respectively, and the bracket represents an annual average over time. For gases like CFC-11 and CFC-12 which have not yet attained their steady-state concentrations in the contemporary atmosphere, one may also define a present state lifetime by the ratio of the current burden to current removal rate[4]. This lifetime is in general time dependent and may differ from the steady-state lifetimes defined by equation (1). (Unless stated otherwise, the term 'lifetime' here will refer to the steady-state lifetime.)

The global burden of N_2O, CFC-11 and CFC-12 may often be estimated from the observed concentrations at a few ground stations because the bulk of these gases reside in the troposphere where they are well mixed[2]. The calculation of the global removal rate is, however, less straightforward as it involves the integral of the product of f, L and N. The major contribution to the integral arises from the stratospheric region where the quantities f, L and N vary significantly with altitudes and latitudes. A common approach is to approximate the integral (which may be interpreted as a spatial and temporal average of the product) by the product of the corresponding averaged quantities. This approach clearly ignores the correlation between the quantities f, L and N and hence its contribution to the integral. The neglect of this correlation may introduce significant errors[5,6] in the calculation of certain reaction rates. We now show that the global removal rate depends not only on the vertical structure of f, L and N in the stratosphere, but also on their latitudinal distributions and correlations.

Three approaches are used to determine the atmospheric lifetimes of the above gases. The first requires empirical knowledge of the source strength and burden. The second approach is to calculate the global removal rate in equation (1) with the spatial distribution of f constructed from available field data, an approach used by Johnston *et al.*[7] in their analysis of N_2O budget. The limitation of these two approaches is that an extensive set of global data is often needed.

In the absence of sufficient data coverage, one must resort to the third approach which uses models to calculate f and τ while using the limited amount of observations to validate the computed gas distributions. A favourable comparison between calculated and observed f may be used as support for the predicted lifetime.

Among the hierarchy of models, the 1-D models are most

0028–0836/82/210317—03$01.00

widely used[1,2]. The problems with 1-D models concern the interpretation of their simulated vertical profiles which are considered to represent some sort of global longitudinal and latitudinal averages. Strictly, 1-D model results should be compared with the observed global averages defined by a large set of spatial and temporal data. Acquisition of such a data set for any trace gas is clearly an enormous task. Even if such global data are available, difficulty still remains in the assessment of the 1-D model predicted lifetimes. Specifically, the global removal rate calculated using the 1-D vertical profiles of f, L and N may still be in error due to the neglect of the zonal and latitudinal correlation between f and L even if the calculated profiles agree with globally averaged observations.

Recent data on N_2O, CFC-11 and CFC-12 (refs 8–14) show that at each given altitude above tropopause, higher concentrations are always found in the tropical stratosphere, a region where the solar UV radiation which dissociates the gases is most intense. Thus, the contribution from the covariance of f and L to the global removal rate could be significant. To account for the covariant effect, multi-dimensional models are needed.

Ideally, a three-dimensional (3-D) model should be used in the lifetime calculation as the spatial and temporal correlations between different quantities (such as L and f) appearing in equation (1) can be properly treated in the integrals. A 3-D study for N_2O has been made by Levy *et al.*[15]. The main disadvantage of using a 3-D model is that it uses a lot of computer time.

The distribution of f calculated from the 2-D zonal-mean model is representative of the zonal average concentrations. If one assumes that the zonal deviation of f is small, the global removal rate can be accurately approximated by integrating the product of the zonal-mean f and zonal-mean L. This assumption is supported by 3-D model simulation of the N_2O distribution[15], and seems to be appropriate for CFC-11 and CFC-12. We shall now use a 2-D model to demonstrate that the latitudinal covariance often exists and may not be negligible for gases such as N_2O, CFC-11 and CFC-12.

Figure 1 shows the N_2O observations[8–14] at different latitudes together with profiles from our 2-D model[16]. Both the data and the model reveal significant vertical as well as latitudinal variations in the N_2O mixing ratio. The overall agreement between model and observation is good.

Figure 2 shows the model calculated column removal rate of N_2O as a function of time and latitude. Note the strong latitudinal gradient, a feature reflecting the important effect of the latitudinal variations in local concentration and local solar flux. During spring and autumn, the maxima occur around the Equator, whereas during summer and winter, they occur about 15° above and below the Equator respectively. Our calculation shows that the net contribution to the global removal rate from the covariance of f and L is positive. The annual average column removal rate for N_2O is 1.2×10^9 molecules cm^{-2} s^{-1}. With a calculated column abundance of 6×10^{18} molecules cm^{-2}, a lifetime of 159 yr is derived. This value may be compared with values of 175 and 150 yr calculated by Johnston *et al.*[7] and Levy *et al.*[15] respectively.

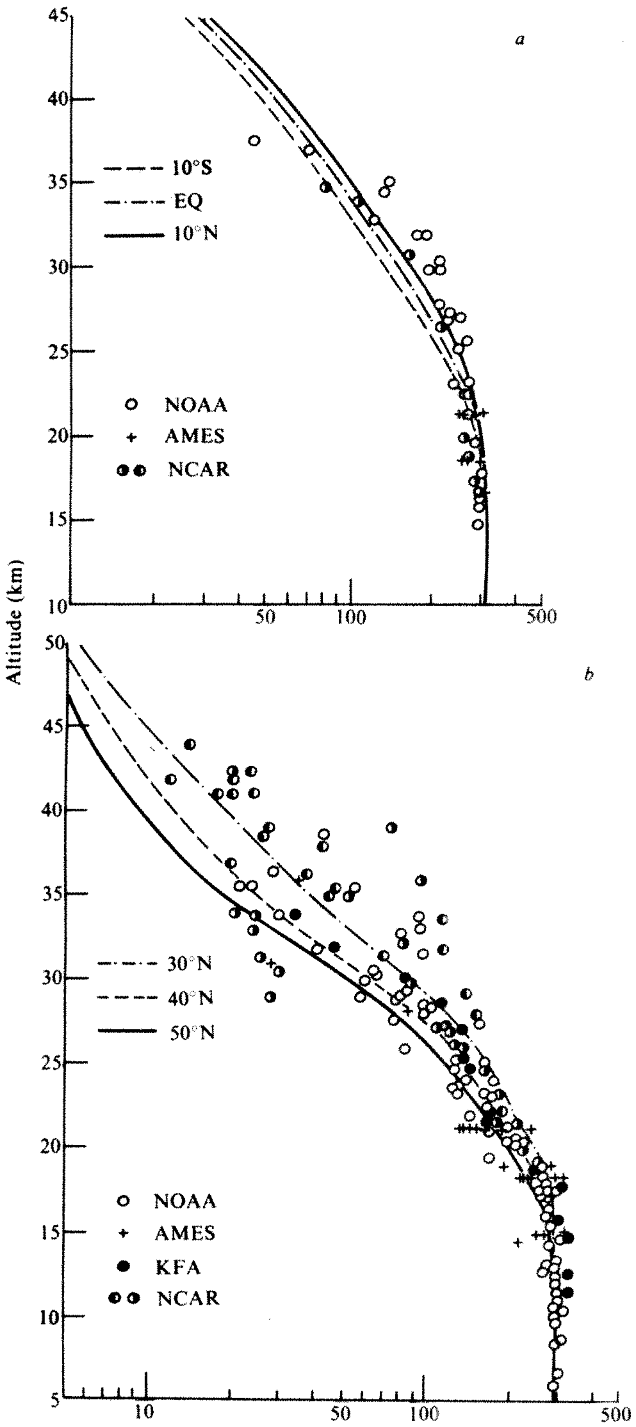


Fig. 1 Calculated N_2O profiles together with observations for equatorial (*a*) and mid-latitude (*b*) regions. The profiles are calculated using the 2-D zonal-mean model described in ref. 16. Dynamic transports in the model are simulated by advection from zonal-mean wind plus eddy transport parameterized by a symmetric eddy diffusion tensor. The wind fields are from ref. 20, while the values of the eddy diffusion tensor are from ref. 21. The reaction rates are from the NASA-WMO Workshop recommendations[18]. A temperature-dependent absorption cross-section[22] is used to calculate the N_2O photolysis rate. A uniform boundary value of 300 p.p.b.v. at the ground is assumed for the calculation. The data shown are reported by groups from NOAA[8,9] NASA–Ames[10–12], KFA[13] and NCAR[14].

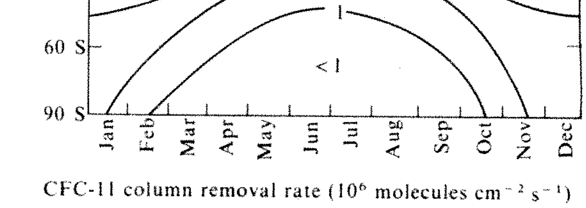


Fig. 2 Calculated N_2O column removal rate as a function of latitude and time of the year.

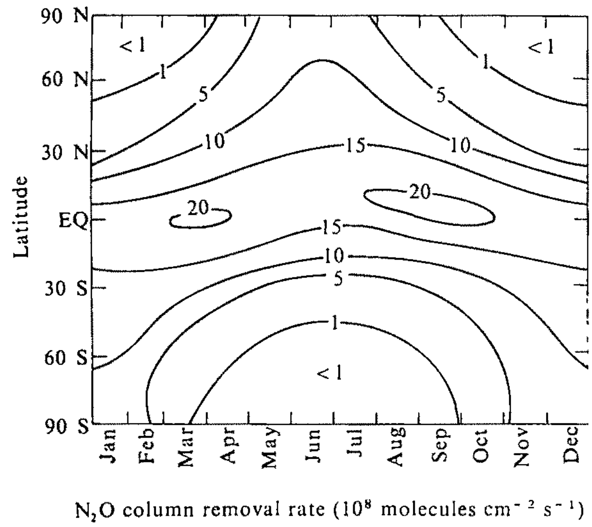


Fig. 3 Calculated CFC-11 column removal rate as a function of latitude and time of the year. The model simulation was performed using a boundary value of 1 p.p.b.v. at the ground.

We have also calculated the lifetimes for CFC-11 and CFC-12 using a steady-state simulation. Direct comparison of the calculated gas distributions of this model simulation with observations is not possible as the gases have not attained the steady-state concentrations in the present atmosphere. As an alternative, the computed CFC-11 and CFC-12 distributions corresponding to the year 1979 of a model simulation using known release rates[17] can be compared with the available observations[8-14]. Agreement is reasonably good for CFC-11 and CFC-12 at the Equator but poor at mid-latitudes[18]. The calculated mid-latitude CFC-11 and CFC-12 mixing ratios above 25 km are higher than the observations by about a factor of 2–4. The discrepancy could be due to either the error in the photolysis rates or error in the transport coefficients. The column removal rate of CFC-12 has features very similar to that of N_2O (Fig. 2). For CFC-11, the latitudinal gradient of the removal rate is stronger as shown in Fig. 3. The calculated lifetimes are 65 and 135 yr for CFC-11 and CFC-12 respectively.

The corresponding 2-D model calculated lifetimes reported by Hinds[19] are 81 and 182 yr and those from the Oxford University model[3] are 78 and 133 yr respectively. Direct comparison of these values with our calculated values is difficult because it is not clear whether the values reported correspond to steady-state lifetimes.

Note that the calculated lifetimes given here also depend on model structure, such as parameterization for transport and treatment of photochemistry. The question of whether these calculated values truly represent the actual atmospheric lifetimes can be answered only following further validation of the corresponding calculated distributions using more observational data and improvement in transport and chemical data. However, we believe that the latitudinal behaviour of the column removal rate simulated by the 2-D model is qualitatively correct and that the positive correlation between the latitudinal variation of f and L should not be ignored.

Although 1-D models are still the most widely used diagnostic and prognostic tools in stratospheric research, two limitations should be noted with regard to their lifetime predictions: first, their inability to incorporate the covariance effects discussed here; second, and more problematic, the inherent difficulty in validating their lifetime predictions using field data. Our discussion points to the fact that, even for a fully validated 1-D model in which the calculated profiles agree with global average observations, the model may still underestimate the global removal rate and, therefore, overestimate the lifetime because of its inability to account for the positive contribution from the covariance of f and L. With a modest increase in computer time, 2-D models, on the other hand, can simulate the latitudinal and seasonal variations of atmospheric species as well as the solar insolation and may therefore be useful in refining the lifetime estimates calculated by 1-D models. In light of the above discussion, a combination of selective field data and a validated 2-D model appears to be the best approach to obtaining more accurate lifetime estimates. Obviously, 2-D models are still in a developmental state, and more atmospheric data are needed for further validation and improvement.

We thank Bruce Lane and Ronald Prinn for useful discussion on the possible difference between present-day and steady-state lifetime, John Pyle for providing the dynamical framework of our zonal-mean model, and P. B. Ryan for analysis of the 2-D model calculations. Research supported by Chemical Manufacturers Association and the Air Force Geophysics Laboratory under contract F19628-81-C-0016.

Received 24 November 1981; accepted 18 March 1982.

1. Schiff, H. I. (ed.) *Stratospheric Ozone: Depletion by Halocarbon, Chemistry and Transport* (National Academy of Sciences, 1979).
2. Hudson, R. D. & Reed, E. I. (eds) *The Stratosphere: Present and Future* (National Aeronautics and Space Adminstration, RP-1049 1979).
3. *Chlorofluorocarbons and their Effects on Stratospheric Ozone* (UK Department of the Environment Pollution Paper No. 15, 1979).
4. Cunnold, D., Alyea, F. N. & Prinn, R. G. *J. geophys. Res.* **83**, 5493–5500 (1978).
5. Tuck, A. F. *Phil. Trans. R. Soc.* **A290**, 477–494 (1979).
6. Derwent, R. G. & Eggleton, A. E. J. *Jl R. met. Soc.* **107**, 231–242 (1981).
7. Johnston, H. S., Serange, O. & Podolske, J. *J. geophys. Res.* **84**, 5077–5082 (1979).
8. Goldan, P. D., Kuster, W. C., Albritton, D. L. & Schmeltekopf, A. L. *J. geophys. Res.* **85**, 413–423 (1980).
9. Goldan, P. D., Kuster, W. C., Schmeltekopf, A. L., Fehsenfeld, F. C. & Albritton, D. L. *J. geophys. Res.* **86**, 5385–5386 (1981).
10. Vedder, J. F., Tyson, B. J., Brewer, R. B., Boitnott, C. A. & Inn, E. C. Y. *Geophys. Res. Lett.* **5**, 33–36 (1978).
11. Tyson, B. J., Vedder, J. F., Arveson, J. C. & Brewer, R. B. *Geophys. Res. Lett.* **5**, 369–372 (1978).
12. Vedder, J. F., Inn, E. C. Y., Tyson, B. J., Boitnott, C. A. & O'Hara, D. *J. geophys. Res.* **86**, 7363–7368 (1981).
13. Volz, A. *et al.* Jul-Report No. 1742 (Kernforschlungsanlage, Jülich, Federal Republic of Germany, 1981).
14. Heidt, L. E., Lueb, R., Pollock, W. & Ehhalt, D. H. *Geophys. Res. Lett.* **2**, 445–447 (1975).
15. Levy, H., Mahlman, J. D. & Moxim, W. J. *Geophys. Res. Lett.* **6**, 155–158 (1979).
16. Ko, M. K. W., Livshits, M. & Sze, N. D. *Proc. Quadrennial Int. Ozone Symp.*, Boulder, Vol. II (ed. London, J.) 884–891 (1980).
17. *World Production and Release of CFC-11 and CFC-12 through 1979* (Chemical Manufacturers Association, 1980).
18. *The Stratosphere 1981: Theory and Measurements* (WMO/NASA, 1981).
19. Hinds, M. K. *Met. Mag.* **108**, 221–240 (1979).
20. Harwood, R. S. & Pyle, J. A. *Q. Jl R. met. Soc.* **103**, 319–343 (1977).
21. Luther, F. M. *AIAA pap.* 73–498. *AIAA/AMS Conf.*, Denver, (1973).
22. Selwyn, G., Podolske, J. & Johnston, H. S. *Geophys. Res. Lett.* **4**, 427–430 (1977).

Position of the Lhasa block, South Tibet, during the late Cretaceous

Jean-Pierre Pozzi*, Michel Westphal†, Yao Xiu Zhou‡, Li Sheng Xing‡ & Xian Yao Chen‡

* Institut de Physique du Globe de Paris, 4 Place Jussieu, F-75230 Paris Cedex 05, France
† Institut de Physique du Globe de Strasbourg, 5 rue Descartes, F-67084 Strasbourg Cedex, France
‡ Chinese Academy of Geological Sciences, Bai Wan Zhuang Road, Beijing, People's Republic of China

The latitude of the Lhasa block (Xizang, South Tibet) during the late Cretaceous–early Tertiary is of prime importance for understanding the collision of India and Eurasia and the mechanism of the formation of the Qinghai–Xizang plateau. During the first field expedition of the French–Chinese collaboration, several geological formations have been sampled in southern Tibet between Lhasa (29.40° N; 91.09° E) and Dingri (28.35° N; 86.38° E). We present here the first results obtained for middle Cretaceous to Palaeocene red beds from the Lhasa block, north of the Yarlung–Zang Bo suture zone, recognized as the main suture between the Indian and the Eurasian plates. They indicate a palaeolatitude of the Lhasa block of 20° N during the late Cretaceous while India was situated 30° S.

0028–0836/82/210319—03$01.00

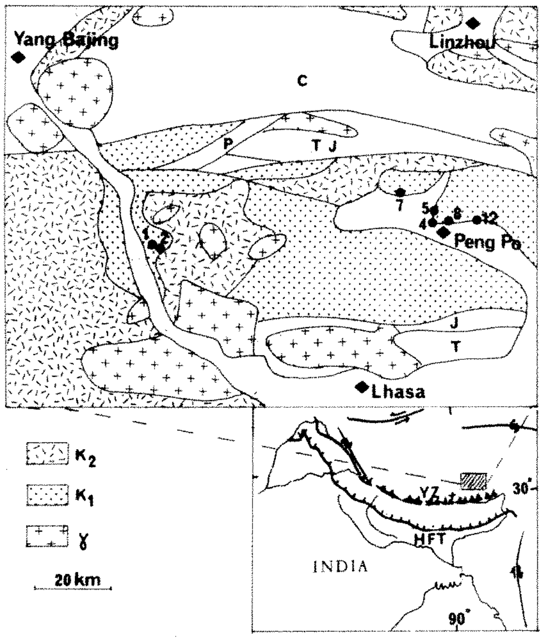


Fig. 1 Geological outlines of the Lhasa region: C, Carboniferous; P, Permian; T, Trias; J, Jurassic. K_1, Takena Formation and lower-middle Cretaceous; K_2, Lingzizong Formation; γ, granitic intrusives. Redrawn from geological map of Lhasa–Nyalam Area (Academia Sinica). YZ, Yarlung Zang Bo suture zone; HFT, Himalayan Frontal Thrust.

The geological formations that we sampled were red beds from Linzhou basin. Continental sediments are present all over the Tibetan plateau and particularly in the Linzhou basin north-east of Lhasa (Fig. 1) where they are represented by two formations separated by a major unconformity. The rock lying under the unconformity is Takena Formation of middle to late Cretaceous age, mainly consisting of sandstone, limestone and shale, strongly folded and with a total thickness of ~1 km. Volcanics (andesite and dacite) are interbedded at the base of this formation. Above the unconformity, Lingzizong Formation of late Cretaceous age mainly consists of black shale in its lower part and red sandstone often conglomerate upwards. A thick sequence of greyish-blue andesite overlays the Lingzizong Formation. Red sandstones of the Takena Formation were sampled (Fig. 1) at sites 1, 2, along the road Lhasa–Yangbajin and at sites 4, 5, 7, 8, 12 near Linzhou (Peng Po Farm).

Samples were taken with a portable drill or as block samples, and were oriented with magnetic and solar compass. Individual sites were sampled laterally over 50–100 m and vertically over 10–20 m.

The measurements were done in Paris and Strasbourg on standard 25-mm cores using either a Schönstedt or a Digico Spinner magnetometer. The two series of measures were cross checked. Samples were demagnetized by step heating and cooling in a zero field (better than 20 nT). The different magnetic components were analysed using the classical Zijderveld vector diagram method.

Thermal cleaning shows the presence of two magnetic components in approximately half of the samples. One is eliminated after heating at 400–500 °C and does not show a clear mean direction. The other is only broken down above 650 °C and is usually the last component. The directions of this component are well grouped. Other samples show only one direction eliminated above 650 °C. Before tectonic correction, the mean directions are generally different from the present field.

Table 1 shows the results before and after tectonic corrections. k increases from 7.3 to 35 by a factor of 4.8. Statistical tests have been used to estimate the validity of the fold test. McElhinny's test[1], which is very stringent, is positive at 95% significance level. McFadden and Jones[2] devised another test but this was difficult to use when each site had a different tectonic setting. We can confidently conclude that the fold test is positive and that the magnetization predates the late Cretaceous–Palaeocene folding phase.

In site 7, four specimens are of normal polarity while seven are inverse, the two directions being antiparallel. The presence of normal and reverse polarity in site 7 of red and green sandstone also argue for a primary origin of the hard component of magnetization.

Nevertheless, against the strong technical arguments for primary origin, one has to consider (Fig. 2) that before bedding correction, the virtual geomagnetic poles (VGP) for sites 1, 2, 4, 5, 8 and 12 are not significantly different (C. T. Klootwijk, personal communication) from the apparent polar wander path (APWP) for India. In the Linzhou basin, tectonic movements of comparatively little amplitude occurred after the main late Cretaceous–early Tertiary folding phase. VGP from site 7 with

Table 1 Palaeogeomagnetic results before and after tectonic correction

	Before unfolding			After unfolding			
Site	N	D	I	D	I	k	$\alpha 95$
Present results							
1	8	335	2	324	37	68	7°
2	7	341	9	333	43	16	16°
4	9	350	42	333	28	90	5°
5	10	352	33	328	47	36	8°
7*	11	10	25	3	24	69	6°
8	11	0	52	338	29	51	6°
12	12	344	−33	341	42	48	6°
Mean before tectonic correction		350	20	—	—	7, 3	24
After tectonic correction		—	—	338	36	35	10
Mean PGV: N = 7 sites Lat. 68° N Long. 340° E							
Palaeolatitude for Lhasa 20° N							
Previous results[3,4]							
Red beds (Linzhou)							
Zhu Xiang Yuan	12	—	—	335	16	66	5°
Zhu Zhi Wen	42	—	—	338	40	20	7°
Lhasa Upper Jurassic	21	—	—	175	2	5	17°

* Both reverse and normal samples.

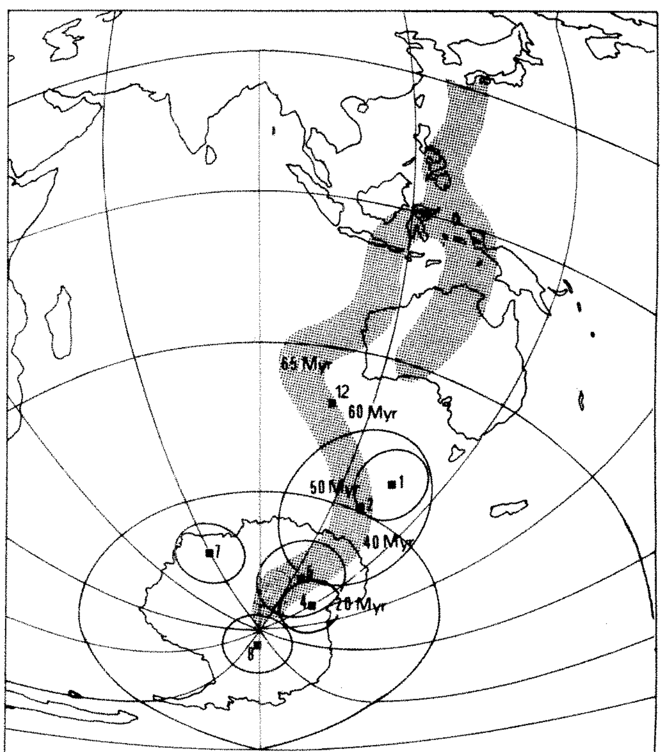


Fig. 2 APWP for India (from ref. 6) compared with the geomagnetic poles of Cretaceous Tibetan sediments and volcanics before unfolding.

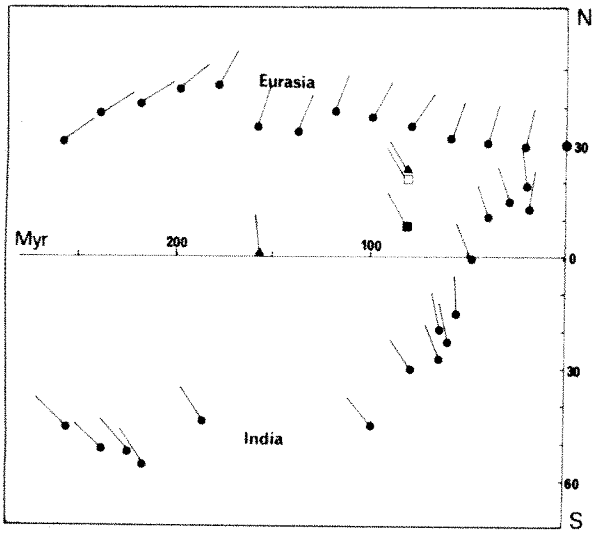


Fig. 3 Palaeolatitude and palaeoorientation of a point now in Lhasa as if it were bound to Eurasia (upper curve), or as if it were bound to India (lower curve). Recalculated from ref. 5 and from refs 6, 7. □, Results from this paper; ▲, results from ref. 4; ■, results from ref. 3.

both normal and reverse layers is different from the APWP of India. Although only a few sites are considered here, it seems that VGPs from site 12, from sites 1 and 2 and from sites 4 and 5 correspond respectively to periods of 65, 50 and 25 Myr on the APWP for India, the two last being critical periods for remagnetization. This could argue for a remagnetization of these formations, when the Lhasa block collided with India. If remagnetization occurred, other methods of investigations will be necessary to decipher the magnetization, or other types of formations should be used to assess the palaeolatitude of Lhasa block; useful conclusions can then be deduced from overprinted components concerning the latitude of the collisions, but it will be difficult to explain how three phases of remagnetization have affected, first site 12, then sites 1 and 2 and possibly sites 4, 5 at 65–50 Myr in a near equatorial latitude, and finally at higher latitude site 4, at 25 Myr.

Until now, the positive fold test has led us to conclude that a complete remagnetization probably did not occur and that the results give a good approximation of the position of the Lhasa block during the Cretaceous.

Chinese palaeomagneticists have done several studies in Xizang, only a few of which have been published. Zhu Xiang Yuan *et al.*[3] have sampled the Takena red beds probably not far from site 4 (the Yayong hill near Peng Po farm). A few samples were a.f. demagnetized up to 30 mT and do not show any directional change. The result given is based on natural remanent magnetism measurements and is dip corrected.

A more extensive study has been published by Zhu Zhi Wen *et al.*[4] over Xizang. Results concerning the Lhasa block are on red beds from Takena (Lhasa cement mill). All samples were a.f. demagnetized up to 50 mT and the results corrected for tilt are given in Table 1.

Compared with the Chinese results for the Linzhou red beds, new results have been obtained on more widely distributed sites using thermal demagnetization techniques which are generally more effective for resolving multi-component magnetizations of red beds than are a.f. techniques. Indeed, a.f. demagnetizations performed on red beds by Zhu Zhi Wen and others leave 90% of the initial magnetization undestroyed at the peak field used (50 mT).

Figure 3 gives the palaeolatitude of Eurasia recalculated from ref. 5 and India from refs 6, 7 as a function of time, with the results of Zhu Zhi Wen[4] for Jurassic limestone and Cretaceous red beds. Arrows indicate declinations with respect to the north and show an anticlockwise rotation of the Lhasa block since the Jurassic. During the middle Cretaceous, the palaeolatitude of the Lhasa block was about 20° N. The results of Zhu Zhi Wen[4] indicate a near equatorial latitude during the Jurassic. These results confirm that the Lhasa block, which is of Gondwanan origin and has preceded northwards drifting India, was widely separated from India during the late Cretaceous and possibly during the Jurassic according to Zhu Zhi Wen results (Fig. 3). During the Cretaceous the northern edge of India was situated near 30° S and the southern margin of Asia probably not far from 30° N. As the palaeolatitude of the Lhasa block was about 20° N it is likely that the Tertiary convergence between South Tibet and India was much larger due to subduction and crustal shortening[8] than between South Tibet and Asia. The anticlockwise rotation of the Lhasa block (Fig. 3) probably occurred under the constraint of the anticlockwise rotation of northward moving India block.

This work was supported by the French CNRS and Institut National d'Astronomie et de Géophysique and by the Ministry of Geology and the Academy of Sciences of the People's Republic of China. Lu Liang Zhong, Yuan Xiang Guo and Ding Fang helped with the field work. We thank Mrs A. Curien for translating several Chinese papers. Contribution IPG no. 586.

Received 26 November 1981; accepted 19 March 1982.

1. McElhinny, M. W. *Geophys. J. astr. Soc.* **8**, 338–430 (1964).
2. McFadden,. P. L. & Jones, D. L. *Geophys. J. R. astr. Soc.* **67**, 53–58 (1981).
3. Yuan, Z. X., Chun, L., Sujuan, Y. & Jinlu, L. *Scient. Geol. Sin.* **1**, 44–51 (1977).
4. Wen, Z. Z., Yuan, Z. X. & Ming, Z. Y. *Acta geophys. Sin.* **24**, 40–49 (1981).
5. Irving, E. *Nature* **170**, 304–309 (1977).
6. Klootwijk, C. T. & Radhakrishnamurty, C. Geodyn. Ser. Vol. 2 (eds McElhinny, M. W. & Valencio, D. A.) (AGU, 1981).
7. Klootwijk, C. T. *Geodynamics of Pakistan* (eds Farah, A. & De Jong, K.) 41–80 (Geological Survey of Pakistan, Quetta, 1979).
8. Tapponnier, P. *et al. Nature* **294**, 405–410 (1981).

Palaeoclimates at Lake Turkana, Kenya, from oxygen isotope ratios of gastropod shells

Paul I. Abell

Chemistry Department, University of Rhode Island, Kingston, Rhode Island 02881, USA

Climatic changes in the Plio-Pleistocene of East Africa during the evolutionary development of early man have been identified only in their broadest outlines. A variety of palaeontological[1–3] and geochemical[4] techniques have already been applied to the study of palaeoclimates in this area, but the use of stable oxygen isotope ratios, so helpful in marine palaeoclimatology, has not been fully exploited[5,6]. The frequent occurrence of well-preserved mollusc shells in the sedimentary column at Lake Turkana provides a sensitive oxygen isotope record, varying over a range of 7‰, which can be interpreted in terms of evaporative concentration of the heavier oxygen isotope in waters of the closed lake, alternating with isotopically lighter freshwater when the lake has been open. I report here the results of a study of the $^{18}O/^{16}O$ ratios of three species of gastropods from the sediments of the Koobi Fora region of Lake Turkana, and evaluate how the technique might be applied to the study of the palaeoclimates of other tropical lakes.

The sedimentary deposits of palaeo-Lake Turkana in northern Kenya are well known for their yield of hominid fossils[7] and artefacts[8] of early man. They also include numerous and varied land and freshwater fossils, including the frequent occurrence of well-preserved mollusc shells. A well-documented collection of these shells[9] provided the impetus and raw materials for this study. Among the 95 shell-bearing horizons recognized by Williamson in his study of the palaeontology of the Lake Turkana molluscs, some 36 contain the gastropods *Mellanoides tuberculata*, *Cleopatra ferruginea* and *Bellamya unicolor*. These three were chosen from among the ~20 species identified by Williamson because: (1) they occur in shallow water habitats and would be expected to be in good equilibrium with the surrounding water; (2) they are capable of living over a wide range of *p*H and concentrations of dissolved salts

0028–0836/82/210321—03$01.00

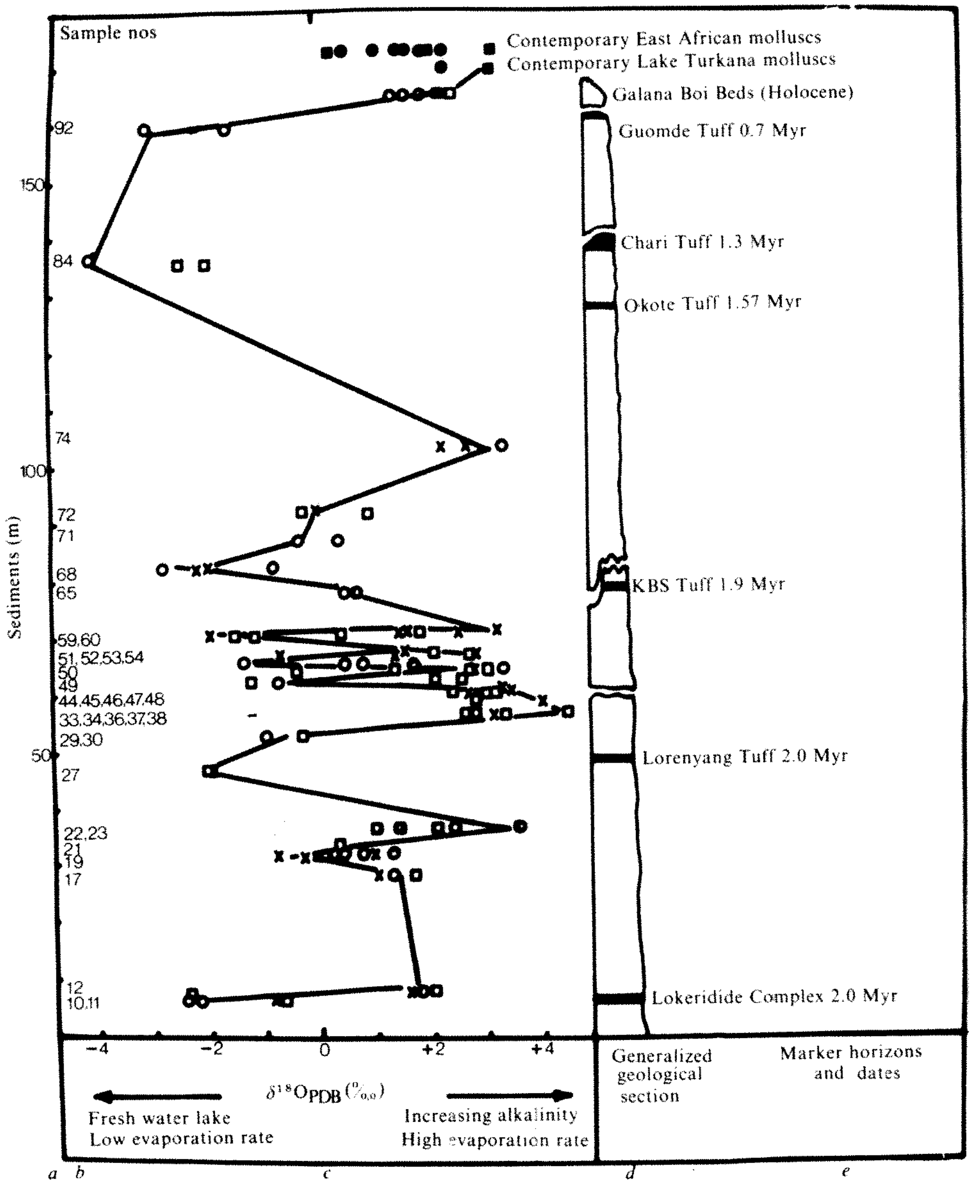


Fig. 1 Variation of $\delta^{18}O_{PDB}$ with stratigraphical level of the fossil molluscs in the Plio-Pleistocene sediments east of Lake Turkana, Kenya. *a*, Sediment depth. *b*, Molluscan stratigraphical levels. These levels were assigned by Williamson[3] and are numbered consecutively from the lowest and oldest horizon where shells are found, upwards. The stratigraphical levels cited include all those in which one or more of the three species of gastropods chosen for this study were found. *c*, $\delta^{18}O$ in ‰ versus the PDB standard. ×, Fossil *Cleopatra ferruginea*; □, fossil *Bellamya unicolor*; ○, fossil *Melanoides tuberculata*, all from the Lake Turkana sediments; ■, contemporary *Bellamya*; ●, contemporary *Melanoides*. In constructing the regression line, generally the most extreme values of $\delta^{18}O$ at each level were selected, representing the extremes of evaporative concentration of $H_2{}^{18}O$, or dilution by meteoric water. This emphasizes short-term extremes of climate over longer-term averages. *d*, Generalized geological section, with disconformities. *e*, Dated marker beds. The contemporary Lake Turkana gastropods came from delta areas near the Omo and Kerio Rivers. The stratigraphy and tuff marker bed correlations are described elsewhere[11,19,20]. Reproducibility for the isotope measurements was ±0.1‰ for multiple analyses of the same individual gastropod and generally better than ±0.5‰ for interspecies comparisons when more than one species was found at the same stratigraphical level.

(although *Bellamya* is more sensitive than the other two); (3) they frequently inhabit the same sedimentary horizon, and thus provide an intergeneric check on the isotope results; and (4) they are the most ubiquitous of the Lake Turkana molluscs, and are thus best suited to give a reasonably continuous climatic record.

Well-preserved specimens were selected and whole shells were ground to a powder and examined by X-ray diffraction for aragonite/calcite ratios. All results reported here are for shells of >90% aragonite. Oxygen isotope ratio measurements were made on a VG Micromass 602C spectrometer, the $\delta^{18}O$ values were calculated in the usual way and are reported here versus the PDB standard[10]. The results are shown in Fig. 1, together with various stratigraphical and sampling data. Details of collection sites, stratigraphy and palaeontology are given elsewhere[3,9,11], while locations and the sample numbers used in this study are recorded in Fig. 2. The results from the Williamson collection were supplemented by a few samples which I collected from the Galana Boi Beds (Holocene), and samples from the modern lake, obtained by dredging near the mouths of the Kerio and Omo Rivers. For comparison contemporary gastropod shells from Lakes Victoria, Edward, Albert and Kivu were analysed and their isotopic results included at the top of Fig. 1. Excursions of the oxygen isotope curve towards positive values represent evaporative regimes and indicate dry and/or hot climates. Excursions towards negative values represent non-evaporative conditions, and indicate cool and/or wet climates.

Note that a few of the climatic fluctuations in Fig. 1 are marked by a substantial range of isotopic values at the same stratigraphic horizon (for example, 59, 60), implying a substantial change in $\delta^{18}O$ with little change in lake level. A rapidly changing microenvironment is the obvious explanation, but no choice can be made among the many detailed scenarios that might produce such changes. These situations are, fortunately, few in number, and do not vitiate the principal trends.

The concentration of the heavier isotope of oxygen in evaporative regimes is well known[12,13] and has been studied over short time spans in both marine and lacustrine environments. The study by Fontes and Gonfiantini[14] on Saharan closed basins, Gara Diba Guelta and Melah Sebkha, is the most pertinent. However, the nearly idealized conditions of their study (limited to 1 yr on a basin with no further input of water or salts during that year) are difficult to extrapolate to Lake Turkana, which has existed for 10,000 yr as a closed basin since its last flushing[15], has had continuous but erratic input of water during that period, and been through a period of considerable volcanic activity by volcanoes actually in the lake, and has been subjected to widely different conditions of relative humidity during those years. Nevertheless, at Lake Turkana, the evaporative history can be deduced approximately because we have three known points defining the relationship between $\delta^{18}O$ and the increasing alkalinity of the lake. This relationship is shown graphically in Fig. 3. In temperate cool/wet climatic conditions, calcium carbonate precipitated from meteoric water has a range of $\delta^{18}O_{PDB}$ at about −5 to −4‰ (ref. 16), depending on the

prevailing temperature[17]. (A 5 °C temperature range, such as currently exists among East African lakes[18], is indicated by the crosshatched area of Fig. 3.) At Lake Turkana, the most positive $\delta^{18}O_{PDB}$ values for mollusc shells fall near +4‰, representing near extinction conditions (16 mequiv. l^{-1} of alkalinity). The present lake, with virtually all molluscs extinct except in areas immediately off the mouths of the Omo, Turkwell and Kerio Rivers, has an alkalinity of 22 mequiv. l^{-1} and a $\delta^{18}O_{SMOW}$ value of +6‰, would be expected to precipitate calcium carbonate with a $\delta^{18}O_{PDB}$ value of nearly 6‰ (ref. 16). Note that Fig. 3 is an approximation for palaeo Lake Turkana only, and a line of drastically different slope could result from similar constructions for other lakes, or even for Lake Turkana in different carbonate supply conditions. (Lake Baringo, $\delta^{18}O_{SMOW}$ = +8.43‰ and an alkalinity of 4.44 mequiv. l^{-1} (ref. 16) represents such a very different lake.)

A similar alkalinity–oxygen isotope curve can be constructed for other palaeolakes, provided, of course, that alkalinity has not accumulated so rapidly that it curtailed molluscan growth early in the history of the lake.

The climatic trends recorded in Fig. 1 for Lake Turkana have some contemporary analogies in Africa. The warm/dry periods which have characterized much of the past at Lake Turkana and which prevail today are also recorded by much the same oxygen isotope ratios in modern gastropods at Lakes Victoria, Edward, Albert and Kivu ($\delta^{18}O_{PDB}$ = 0 to +4‰), suggesting that temperature more than relative humidity controls the

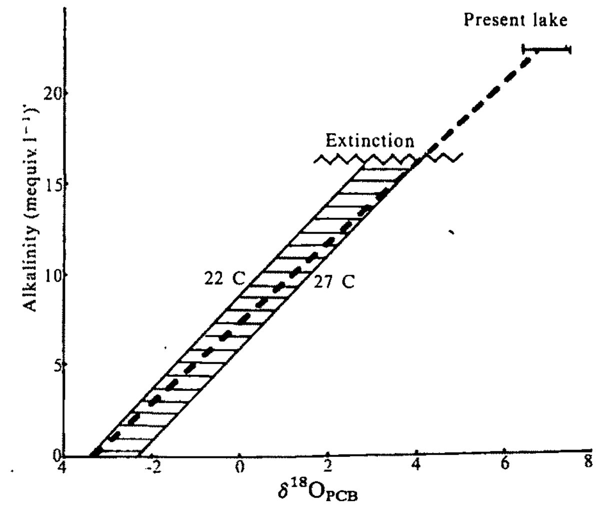


Fig. 3 Change in $\delta^{18}O_{PDB}$ of gastropods at Lake Turkana as a function of progressive increase in alkalinity. Meteoric waters, subjected to little evaporative loss, would yield molluscs of $\delta^{18}O_{PDB} \cong -4‰$ at near zero alkalinity. Progressive evaporation of the lake in a closed basin would increase the alkalinity and the $\delta^{18}O$ of the molluscs along the curve to the right. Extinction occurs at an alkalinity of about 16 mequiv. l^{-1}. The cross-hatched area represents temperature effects on $\delta^{18}O$ (ref. 17), assuming the normal temperature range of contemporary East African lakes of ~5 °C (ref. 18).

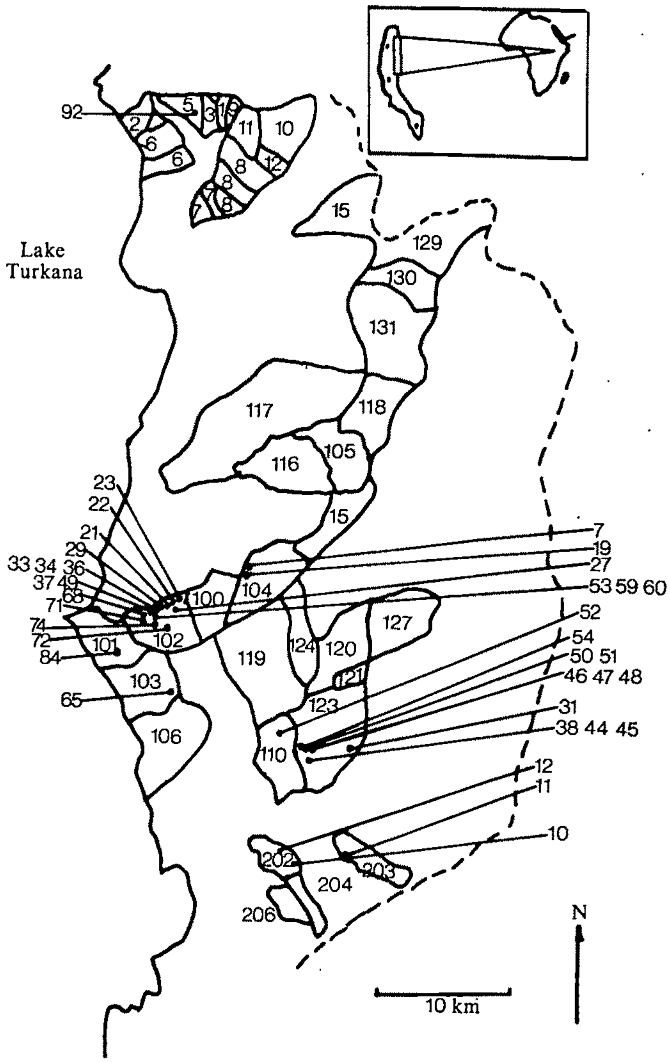


Fig. 2 Koobi Fora region of the Lake Turkana Basin, Kenya, showing the sample numbers (molluscan stratigraphic horizons) and collection sites of the gastropods used in this study in the topographically delineated fossil localities[3].

evaporative concentration of ^{18}O. On the other hand, the negative values of $\delta^{18}O$ which have been observed for gastropods of historical times are similar to the isotope ratios of contemporary gastropods of rivers in the Transvaal ($\delta^{18}O_{PDB}$ = −2 to −4‰) or in southern Zaire (−3.5‰). Lake Turkana thus had quite a temperate climate at times in the past.

It is unfortunate that the only dated sediments at Lake Turkana are the tuff horizons, so that correlations with the well-dated marine climatic fluctuations cannot be very precise. However, the more important use for this technique is to secure climatic information to be associated with particular fossil beds or archaeological sites, and thus define the climatic component of the evolutionary pressures on the animal population, including man.

I thank Peter Williamson for providing the molluscs for this study, Tom Rockett for assistance in the X-ray analyses, Cecile Vanech for assistance in preparation of the carbon dioxide samples, the Graduate School of Oceanography at this university for use of the mass spectrometer, Andrew Cohen for furnishing samples of contemporary molluscs from Lake Turkana, the Museum of Comparative Zoology of Harvard University for providing mollusc samples from many African lakes, and particularly Richard Leakey for field support at Lake Turkana.

Received 30 November 1981; accepted 19 March 1982.

1. Bonnefille, R. *Bull. Soc. geol. Fr.* **21,** 331–342 (1979).
2. Harris, J. M. in *Earliest Man and Environments in the Lake Turkana Basin* (eds Coppens, Y. *et al.*) 293–301 (University of Chicago Press, 1976).
3. Williamson, P. G. *Nature* **293,** 437–443 (1981).
4. Cerling, T. E., Hay, R. L. & O'Neil, J. R. *Nature* **267,** 137–138 (1977).
5. Cerling, T. E. *Palaeogeogr., Palaeoclimatol., Palaeoecol.* **27,** 247–285 (1979).
6. Abell, P. I., Awramik, S. M., Osborne, R. H. & Tomellini, S. *Sed. Geol.* (in the press).
7. Leakey, R. E. & Leakey, M. *Koobi Fora Research Project Monogr.* Vol. 1 (Clarendon, Oxford, 1979).
8. Isaac, G. L. in *Earliest Man and Environments in the Lake Turkana Basin* (eds Coppens, Y. *et al.*) 533–551 (University of Chicago Press, 1976).
9. Williamson, P. G. thesis, Univ. Bristol (1980).
10. Craig, H. *Geochim. cosmochim. Acta* **3,** 53–92 (1953).
11. Williamson, P. G. *Nature* **295,** 140–142 (1982).
12. Craig, H. *Science* **133,** 1702–1703 (1961).
13. Lloyd, R. M. *Geochim. cosmochim. Acta* **30,** 801–814 (1966).
14. Fontes, J.-C. & Gonfiantini, R. *Earth planet. Sci. Lett.* **3,** 258–266 (1967).
15. Butzer, K. W., Isaac, G. L., Richardson, J. L. & Washbourn-Kamau, C. *Science* **175,** 1069–1076 (1972).
16. Cerling, T. E. thesis, Univ. California, Berkeley (1977).
17. Craig, H. in *Conf. Stable Isotopes in Oceanographic Studies and Paleotemperatures,* 3–24 (1965).
18. Beadle, L. C. *The Inland Waters of Tropical Africa* (Longmans, London, 1974).
19. Cerling, T. E. & Brown, F. H. *Nature* (in the press).
20. Brown, F. H. & Cerling, T. E. *Nature* (in the press).

Reconstruction of spatial information in the human visual system

Göte Nyman & Pentti Laurinen

Department of General Psychology, University of Helsiniki, Ritarikatu 5, 00170 Helsinki 17, Finland

The remarkable capacity of an observer to perceive and recognize objects and forms independent of the exact nature of their components can be described as a process of reconstruction. We have measured the ability of human observers to recognize square-wave or sinusoidal gratings when presented as a pattern of regularly taken spatial samples (see Fig. 1). To recognize the waveforms, the observer must visually reconstruct them from the samples—a process which can be described by the Shannon–Whittaker theorem[1,2] of sampling and which Barlow suggests may be carried out by the numerous stellate cells of the visual cortex[3]. We report here that at low spatial frequencies, relatively more sample lines per spatial cycle were needed for wave recognition than were needed at higher frequencies. However, a square-wave grating was still recognized easily even when it was sampled at a rate at which its third harmonic could not be recognized when presented alone. At high spatial frequencies the square wave was identified when the sampling rate was so low that it caused the third harmonic to be undersampled. This contradicts the idea that a complex wave is analysed by parallel spatial frequency channels[4,5] and emphasizes the capacity of the visual system to use signal features other than the harmonic frequency components of an image to recognize it.

From the Shannon–Whittaker sampling theorem[1,2] we know that a signal having a limited bandwidth B can be completely reconstructed from samples provided that the samples are taken densely enough, that is, above the minimum sampling frequency of $2B$ (the Nyquist frequency). The signal can be reconstructed by suitable interpolation between the samples or, conversely, by reconstruction filtering. In the psychophysical context described here, the term 'reconstruct' refers to the observer's ability to perceive certain predetermined spatial features, such as waveform or spatial frequency, in a stimulus consisting of spatial samples of a grating of particular waveform and spatial frequency. The sampling theorem cannot be applied to spatial vision in a straightforward manner, because the number of samples per cycle needed for signal reconstruction actually depends on the spatial frequency value of the grating (that is, the number of cycles $\deg^{-1}$ the grating subtends at the observer's eye). When viewed close to, the coarsely sampled gratings on the right and left in Fig. 1 cannot be recognized as sinusoid and square wave, respectively, but at far enough distance the 'grating Gestalts' become visible in the sampled images.

We have measured psychophysically how many samples per spatial cycle are needed to recognize sinusoidal and square-wave gratings. Two gratings were presented successively for 600 ms. The first was always a 'sampled' grating consisting of vertical sample lines of width 4 min. The samples were taken either from a sinusoid with contrast $c = 0.80$ or from a square wave with $c = 0.63$, where $c = (L_{max} - L_{min})/(L_{max} + L_{min})$ and L_{max} and L_{min} refer to the maximum and minimum luminances, respectively, in the grating. With these contrasts the fundamental frequency component of the square wave and the sinusoid of corresponding spatial frequency both have the same energy content. The samples were taken at regular intervals and presented against a background of constant mean luminance, 120 cd m^{-2}. The second stimulus was a continuous grating, also vertical, but with lower contrast (0.50 in sinusoids and 0.31 in square waves). This made the perceived contrast of the reference and the sampled grating approximately similar. The subject had to indicate whether or not the appearance of the first grating, the sampled one, matched the reference grating in certain required properties. The density of sampling was increased or decreased according to the response, to determine the subject's threshold. Dense sampling results in a clearly visible grating of continuous appearance, when the sampling rate is decreased, the sample lines become visible, but a total 'grating Gesalt' can still be seen.

Two different threshold criteria were used. First, the subject decided whether he could see the same spatial frequency and waveform in both of the gratings. If he could not, the sampling rate was increased, otherwise it was decreased. Then he indicated when he was sure that no features of the reference grating waveform remained in the sampled image. This was done in order to estimate how much the criterion level shifts affected the results. The thresholds were measured separately for the sinusoids and the square waves and four spatial frequency values were used. The threshold sampling rate was calculated as the mean of the 10 last minium and maximum values occurring during a threshold run.

The results are expressed in terms of sampling efficiency, which relates the measured threshold sampling rate to the Shannon–Whittaker or Nyquist limit of the grating frequency which was studied. The sampling efficiency $E(s) = N(s)/S(s)$, where s is the spatial frequency, $N(s)$ is the corresponding Nyquist rate and $S(s)$ is the measured value. Thus, $E(s)$ describes the functional sampling efficiency of the visual system in recognition tasks which require spatial interpolation or reconstruction. The results in Fig. 2 indicate that almost identical $E(s)$ values were obtained for the two different waveforms. For both waveforms, $E(s)$ increased as a function of spatial frequency. When the spatial phase of the grating at which the sampling started had the value 0, $\pi/4$, $\pi/2$, $3\pi/4$ or π, similar results were obtained, suggesting that the phase was not critical. This was not surprising because the sampling efficiency was rather low at the spatial frequencies used, at 8 $c \deg^{-1}$ being still

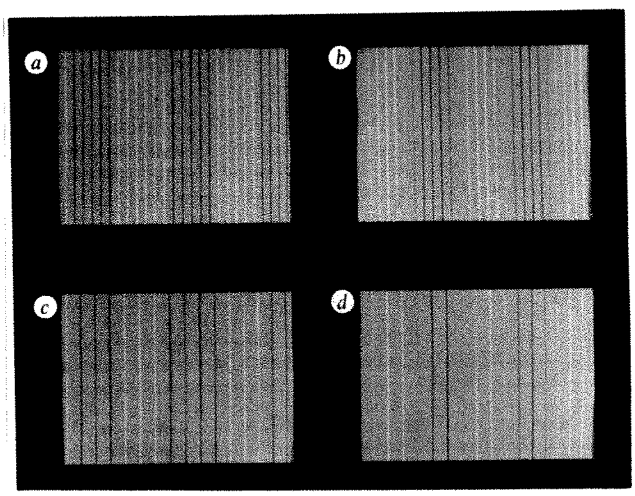


Fig. 1 Spatial sampling effects in grating perception. Square-wave (a, c) and sinusoidal (b, d) gratings of equal spatial frequency have been sampled at regular intervals. Only the sample values carry the grating information and the spaces between them have a constant background value. a, b, Sampling rate of 11 lines per cycle. Increasing the viewing distance (that is, the spatial frequency of the gratings and the angular sampling density) makes it easier to identify the two waveforms even though the physical sampling constraints remain the same. c, d, The same gratings as in a and b, but with sampling density of 7 lines per cycle. The waveforms can be recognized only by increasing the viewing distance sufficiently. The edges of the square wave (a, c) are seen first and it does not have the appearance of its fundamental frequency component. Furthermore, both waveforms are identified simultaneously when the viewing distance is increased. Quantitative results of the sampling and reconstruction effects ar presented in Fig, 2.

0028–0836/82/210324—02$01.00

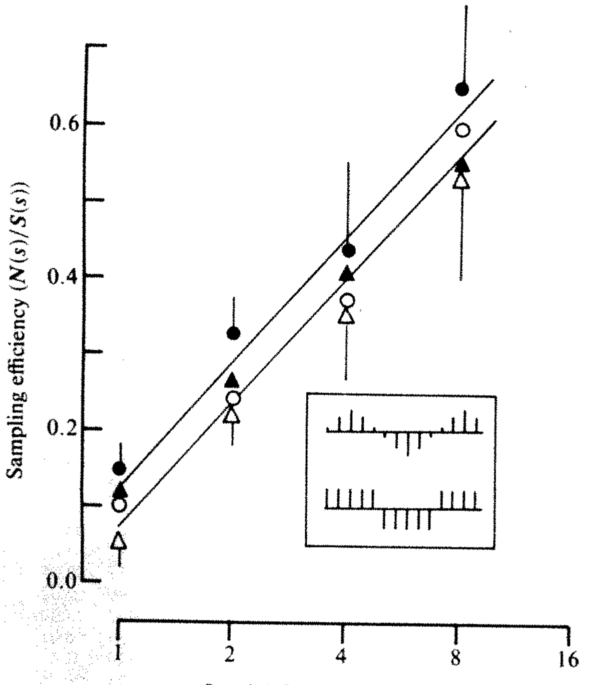


Fig. 2 Sampling efficiency of human spatial vision. We studied the number of sample liners per cycle needed for grating waveform recognition . Two different waveforms were used: ▲, △, sinusoid; ●, ○, square wave. The stimuli from which the samples were taken had contrasts of 0.80 and 0.63, respectively, which makes the energy content of the sinusoid equal to that of the fundamental component in the square wave having the same spatial frequency. The subject adjusted the density of sampling in the grating until he decided that it had the required appearance. Two criteria for the appearance were used: threshold for recognizing the waveform (open symbols) and threshold for the disappearance of the recognizable waveform (filled symbols). The threshold value was taken as the mean of the last 10 settings. The results are expressed in terms of sampling efficiency, $E(s) = N(s)/S(s)$, Where $N(s)$ is the Nyquist sampling frequency at spatial frequency s and $S(s)$ is the threshold value obtained. Viewing was monocular, with natural pupil. Mean luminance was 120 cd m^{-2}, the grating size 28 × 21 cm, and viewing distance 798 cm. Results for one subject (P. L.), an experienced and emmetropic observer, are shown. Experiments with three other subjects gave similar results. The bars represent 1 s.d. and for clarity, only some representative s.d. values are shown. The lines have been drawn by eye and indicate the results for the two different waveforms. The inset describes the waveforms sampled.

~0.6. Changing the threshold criteria caused an average change of 0.10 in efficiency.

If the sampled square wave is recognized when enough of its harmonics contribute to its appearance, then it follows that the sampling rate needed for its recognition should be high enough to make these harmonics recognizable when presented alone. In contrast, the results show that the square wave was recognized even if it was sampled at a rate at which the third harmonic could not be seen. Furthermore, for higher spatial frequencies, the threshold sampling rates for identifying the square wave grating were so low that an efficiency value better than 0.33 was achieved, indicating that the third harmonic of the image must have been undersampled.

Our results contradict the models which describe the visual system as a Fourier analyzer in which a complex waveform is recognized on the basis of the information obtained from its harmonics, each of which is channelled through its own spatial frequency selective channel[4,5]. The visual system can also use other features than the harmonic frequency components of an image to recognize it. In the case of a high frequency square-wave grating it performs better than expected on the basis of the sampling theorem. Such performance is known to be possible in some artificial signal processing systems[6,7] and has also been demonstrated for a model of cortical information processing[8,9].

Received 21 September 1981; accepted 1 April 1982.

1. Shannon, C. E. *Proc. IRE* **37**, 10 (1949).
2. Whittaker, E. T. *Proc. R. Soc. Edinb.* **A35**, 181 (1915).
3. Barlow, H. B. *Nature* **279**, 189 (1979).
4. Campbell, F. W. & Robson, J. G. *J. Physiol., Lond.* **197**, 551 (1968).
5. Maffei, L. & Fiorentini, A. *Nature* **240**, 479 (1972).
6. Logan, B. F. *Bell Syst. tech. J.* **56**, 487 (1977).
7. Bracewell, R. *The Fourier Transform and Its Applications* (McGraw-Hill, New York, 1978).
8. Marr, D. *Phil. Trans. R. Soc.* **B275**, 483 (1976).
9. Marr, D. & Hildreth, E. *Proc. R. Soc.* **B207**, 187 (1980).

Local cerebral glucose utilization in non-rapid eye movement sleep

C. Kennedy*†, J. C. Gillin‡, W. Mendelson‡, S. Suda*, M. Miyaoka*, M. Ito*, R. K. Nakamura§, F. I. Storch‡, K. Pettigrew‖, M. Mishkin§ & L. Sokoloff*

* Laboratory of Cerebral Metabolism, ‡ Biological and Adult Psychiatry Branches, § Laboratory of Neuropsychology, and ‖ Theoretical Statistics and Mathematics Branch, National Institute of Mental Health, US Public Health Service, Department of Health and Human Services, Bethesda, Maryland 20205, USA
† Department of Pediatrics, Georgetown University School of Medicine, Washington, DC, USA

Sleep is accompanied by alterations in neurophysiological activity in many discrete regions in the brain[1]. Such alterations could be expected to be accompanied by corresponding changes in local metabolic rate[2]. We have now looked for the specific regions involved in sleep, at least in the stages other than rapid eye movement (REM) sleep, by using the [2-^{14}C]deoxyglucose method for measuring local cerebral glucose utilization[3] in the rhesus monkey. Contrary to what many have predicted, animals during stages 2–4 of sleep exhibited a non-selective, generalized 30% decrease in cerebral metabolic rate. Of the 75 structures measured, none exhibited a higher rate in non-REM sleep than in wakefulness.

The study was carried out in four awake and four sleeping rhesus monkeys weighing 3.8–5.9 kg. They were surgically prepared several weeks before the experiment with electrodes implanted over the dura, in the orbital region and in neck muscles to permit electrical monitoring of cerebral, ocular and muscular activities. The animals spent daytime hours in cages in which they were free to move. To accustom them to the experimental conditions, we placed them in a restraining chair at a regular hour in the early evening. The chair was moved to a ventilated, darkened chamber in which white noise was continuously generated at an intensity of 70 dB to mask adventitious sounds outside the chamber. The animals readily accepted this environment and would sleep normally each night.

On the day of the experiment catheters were inserted in the femoral artery and vein of one leg to allow the intravenous administration of ^{14}C-deoxyglucose and arterial sampling. This was done under light halothane anaesthesia, from which they took 5–10 min to recover. Three hours later the animals were placed in the sleep chamber. The experimental procedure was initiated in the four sleeping animals ~5 min into the period of slow-wave sleep that immediately followed the first period of REM sleep. Three of the control (awake) animals were also allowed to fall asleep. Five minutes after their first period of REM sleep, however, they were aroused by a gentle jostling of the chamber after which the experiment was begun. The jostling was occasionally repeated when necessary to maintain

wakefulness. A fourth awake control was studied before the onset of sleep. In addition to polygraphic monitoring, a video camera in IR light was used to assess the animal's degree of consciousness during the experiment. In the four sleeping animals, on the average 23 min of the 30 min experimental period was occupied by sleep stages 3 and 4. There were occasional, randomly distributed lapses into stage 2 which lasted on the average 5 min. These lapses took place in only two monkeys during the crucial first 15 min and both lasted 1.5–2 min. Periods of REM sleep and wakefulness, averaging 1 min each, occurred during the last 10 min, when their effect on the results would be negligible due to the relatively low concentration of deoxyglucose in the blood at that time.

The experiments were initiated by an intravenous pulse of 2-deoxy-D-[1-^{14}C]glucose (specific activity 50–56 mCi mmol^{-1}) dissolved in physiological saline in a dose of 100 μCi per kg. The procedure was carried out as previously described[4]. For each animal, approximately 400 brain sections in the Horsley–Clarke frontal plane were prepared for quantitative autoradiography together with a smaller number for histological staining. The local rates of glucose utilization were calculated from 12 measurements of optical density for each of 75 structures in each animal, by means of a transmission densitometer. The aperture used was 0.5 or 0.25 mm, depending on the size of the structure. When the location of a small nucleus in the autoradiograph was uncertain, positive identification was made with the aid of an adjacent histological section stained with cresyl violet.

Midway through the procedure, arterial blood was sampled for the determination of *p*H and the partial pressure of the respiratory gases. The mean values ±s.e.m. of 7.47±0.01,

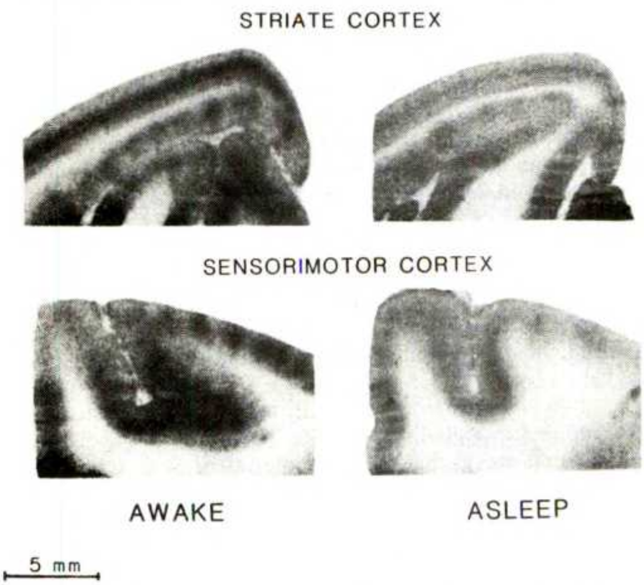


Fig. 1 ^{14}C-deoxyglucose autoradiographs of coronal sections of striate/prestriate cortex (above) and sensorimotor cortex (below) of monkey brain. The sections on the left are representative of the group of awake monkeys, those on the right, of the group asleep. In the striate/prestriate cortex of the awake monkey, the dark bifid line parallel to the cortical surface corresponds to Layer IV. Fine, regularly spaced markings perpendicular to the cortical surface traverse the entire cortical thickness. Directly beneath the cortex on the upper surface and parallel to it is part of the prestriate cortex. Here Layer IV is not seen but there are rather broad perpendicular markings which are irregularly spaced. In the corresponding section of striate/prestriate cortex of a sleeping animal (upper right), these markings are virtually indistinguishable. The narrow dark and light bands (lower right hand side of the section from the sleeping animal) are artefacts due to wrinkling of the section at the time of cutting. In the sensorimotor cortex of the awake animal (lower left), both perpendicular and horizontal markings are seen. In sleep (right), all these markings virtually disappear in both cortical regions, and the metabolic rate is nearly uniform in the various cortical layers.

Table 1 Glucose utilization in monkey brain during slow-wave sleep (μmol per 100 g per min)

Structure	Awake ($n=4$)	Asleep ($n=4$)	% Change	*P* value
Cerebral cortex				
Striate	46±3	30±5	−35	<0.05
Inferior parietal	44±2	28±2	−38	<0.01
Auditory	81±2	46±0.1	−44	<0.0001
Accessory auditory	45±3	36±2	−21	<0.05
Somatosensory	59±8	42±2	−28	<0.1>0.05
Motor	44±3	32±2	−26	<0.05
Inferior temporal	48±3	40±3	−16	0.1
Cingulate	51±4	35±4	−30	<0.05
Perirhinal	27±2	22±2	−21	<0.2
Prefrontal	59±6	43±3	−28	<0.05
Medial orbital	51±4	34±1	−17	<0.01
Thalamus				
Lateral geniculate body	43±1	25±1	−41	<0.0001
Medial geniculate body	73±2	46±1	−37	<0.0001
Dorsal medial n.	53±3	34±2	−37	0.001
Ventral post. lat. n.	43±3	30±1	−29	<0.01
Lateral dorsal n.	43±3	27±2	−37	<0.01
Central lateral n.	44±2	30±6	−33	0.05
Midline n.	37±4	28±2	−23	0.10
Anterior n.	39±1	33±4	−17	0.2
Anterior medial n.	52±6	34±2	−33	<0.05
Reticular n.	28±3	17±1	−40	<0.02
Pulvinar	43±1	29±2	−32	<0.001
Ventral anterior n.	39±2	26±2	−34	<0.01
Lateral habenula	48±3	41±3	−14	0.2
Medial habenula	35±2	28±2	−21	<0.05
Hypothalamus				
Ventromedial n.	26±1	21±2	−20	<0.1>0.05
Supraoptic n.	29±1	22±2	−23	<0.02
Suprachiasmatic n.	28±2	22±1	−21	<0.05
Medial mamillary n.	63±4	47±1	−26	0.01
Preoptic area	27±1	21±2	−22	<0.05
Lateral hypothalamic n.	26±1	21±2	−18	<0.1>0.05
Limbic system				
Lateral amygdaloid n.	31±3	25±2	−20	0.1
Basal amygdaloid n.	27±2	22±1	−16	0.2
Hippocampus	34±3	28±2	−19	<0.1>0.05
Hippocampal gyrus	39±3	30±3	−24	<0.1>0.05
Dentate gyrus	34±3	28±2	−16	0.2
Lateral septal n.	22±2	20±2	−9	0.5
Medial septal n.	29±2	21±1	−29	<0.01
Substantia innominata	26±2	18±1	−29	<0.01
Extrapyramidal motor system				
Caudate n.	59±5	44±2	−26	<0.05
Putamen	59±4	49±1	−18	<0.05
Globus pallidus	24±2	20±1	−15	0.1
Nucleus accumbens	35±4	29±2	−15	0.3
Claustrum	27±2	24±2	−11	0.4
Subthalamus n.	62±5	40±2	−35	<0.01
Substantia nigra	34±3	24±1	−30	0.02
Red nucleus	49±3	36±1	−27	<0.01
Brain stem and midbrain				
Reticular mesencephalic n.	33±1	21±2	−36	<0.001
Oculomotor n.	59±3	46±2	−23	0.01
Interpeduncular n.	41±2	38±3	−6	0.6
Superior olive	93±14	68±5	−27	0.1
Raphe n., dorsal	34±3	29±2	−16	0.2
Raphe n., ventral	35±4	28±2	−22	0.1
Reticular n., magnocellular	34±3	25±2	−26	<0.05
Locus coeruleus	33±3	27±3	−19	0.2
Superior colliculus	55±2	37±2	−33	<0.001
Inferior colliculus	160±10	126±13	−21	<0.1>0.05
Central gray substance	36±1	27±3	−25	<0.02
Ventral tegmental n.	37±2	28±2	−23	<0.05
Vestibular n.	78±8	59±5	−25	<0.1>0.05
Cochlear n.	67±1	64±2	−5	0.1
Caud. mesenceph. retic. n.	34±1	24±3	−29	<0.02
Solitary tract, n.	38±5	32±4	−15	0.4
Area postrema	31±0.6	20±2	−36	0.001
Cerebellum				
Crus I	34±2	24±3	−30	<0.05
Vermis	36±2	24±1	−33	<0.01
Flocculus	53±3	28±2	−46	<0.001
Dentate n.	61±6	49±3	−19	0.1
White matter				
Corpus callosum	13±0.7	10±1	−20	<0.05
Anterior commissure	10±0.3	10±0.5	−7	0.3
Internal capsule	14±1	12±0.9	−15	0.3
Corona radiata	8±0.6	5±0.4	−38	<0.01
Optic tract	18±4	14±0.3	−21	0.4
Fasciculus retroflexus	32±1	24±2	−23	<0.02
Pyramids	10±2	7±1	−36	0.1

Values are mean ±s.e.m.

40 ± 2, 88 ± 2 mm Hg for pH, P_{CO_2} and P_{O_2}, respectively, in awake animals were not significantly different from those of 7.43 ± 0.02, 46 ± 2 and 91 ± 3 in sleep.

The rates of local cerebral glucose utilization in the awake controls in this study were similar to those previously obtained by Kennedy *et al.*[4] in awake animals in the laboratory, except for the rates in a few structures of the auditory and visual pathways. The alterations in these structures could be accounted for by the differing environmental conditions.

In the present study, mean rates of glucose utilization were lower in sleeping monkeys than in awake controls in each of the 75 structures examined, and the difference was significant ($P < 0.05$) in 44 cases (Table 1). The reductions were largest and most consistent in cerebral cortex, cerebellum and thalamus, but no anatomic or functional system was selectively affected, and in no structure was the value in sleep above that during wakefulness. There is, therefore, no evidence to suggest that any cerebral centres were activated during slow-wave sleep. Rates of glucose utilization in the cerebral cortex of the awake animals were seen to be variable in the autoradiographs, which revealed dark markings perpendicular to, and in some areas also parallel to, the cortical surface. Such markings were found in nearly all the major subdivisions of the cortex and, in some cases, the pattern was characteristic for the given area, such as the striate cortex (Fig. 1). This variable pattern of metabolic rates, reflecting the columnar organization of cortical function, virtually disappeared from all regions of the cortex in sleeping animals (Fig. 1).

It has recently been reported that the rate of glucose utilization in the choroid plexus of the cat is elevated in slow-wave sleep[5]. Although the kinetics of deoxyglucose in this tissue has not been sufficiently analysed to ensure accurate quantification of its rate of glucose utilization, the calculation used in this study for brain tissue probably provide at least an index of its metabolic rate. We used a computerized image-processing system[6] to obtain values for choroid plexus in the monkeys. The brain sections containing this tissue were enlarged 15 times to display the irregular borders which then could be traced with an outline marker. Thus it was possible to obtain an integrated value for the structure in each section. Values did not vary significantly in the lateral, third and fourth ventricles, and no statistically significant differences were found between sleeping and waking animals.

It was possible to calculate a single weighted average rate of glucose utilization of the brain as a whole by scanning all sections with the computerized image-processing system previously described[6]. The mean ±s.e.m. was 35 ± 2 μmol per 100 g per min for the four awake control animals. This value was not statistically significantly different from the value of 36 ± 1 μmol per 100 g per min found by Kennedy *et al.* for the previously published results from seven conscious monkeys studied in the ambient environment of the laboratory[4]. Thus the awake control animals in the present study did not have a generalized activation of cerebral metabolic rate due to the experimental conditions, such as the confinement in a closed chamber and the acoustic stimulation. In the four sleeping monkeys, the average rate of cerebral glucose utilization was 25 ± 1 μmol per 100 g per min, a statistically significant reduction ($P < 0.02$) below either of the awake groups.

The results of this study might seem inconsistent with those of Mangold *et al.*[7], who used the Kety–Schmidt method and found no difference in cerebral metabolic rate between slow-wave sleep and wakefulness in human subjects. In five subjects, cerebral oxygen consumption declined during sleep, but in the remaining subject it increased by 34%. As a result, in the group as a whole, cerebral oxygen consumption was not statistically significantly different in sleep and wakefulness. The large difference in the one subject, however, indicates that this subject is a statistical outlier by the criteria of either Grubbs[8] or Dixon[9]. Thus there is a statistical rationale for his elimination from the series. Mangold *et al.*[7] had no reason to discard the findings on the basis of the subject's behaviour or any aspect of the procedure, but had they done so for the statistical reasons cited, the mean value for cerebral oxygen consumption in the remaining five subjects would have declined during sleep by a statistically significant ($P < 0.03$) 11%. The earlier study in man would then be qualitatively consistent with the findings reported in the present studies in the monkey. The discrepancy in the magnitude of the effects may be due to differences in the methods, species, or conditions in which the studies were conducted. In the present study, the measurements were made at the usually normal sleeping time of the monkeys; in the earlier studies in man, the measurements were made at 4:00–6:00 a.m. following deprivation of sleep at the normal scheduled time, and the individuals could have been fatigued.

Received 24 August 1981; accepted 5 April 1982.

1. Steriade, M. & Hobson, J. A. *Prog. Neurobiol.* **6**, 155–376 (1976).
2. Sokoloff, L. *J. Neurochem.* **29**, 13–26 (1977).
3. Sokoloff, L. *et al. J. Neurochem.* **28**, 897–916 (1977).
4. Kennedy, C., Sakurada, O., Shinohara, M., Jehle, J. & Sokoloff, L. *Ann. Neurol.* **4**, 293–301 (1978).
5. Bobillier, P. *et al. C.r. hebd. Séanc. Acad. Sci., Paris* **291**, 91–96 (1980).
6. Goochee, C., Rasband, W. & Sokoloff, L. *Ann. Neurol.* **7**, 359–370 (1980).
7. Mangold, R., Sokoloff, L., Conner, E., Kleinerman, J., Therman, P. G. & Kety, S. S. *J. clin. Invest.* **34**, 1092–1100 (1955).
8. Grubbs, F. E. *Technometrics* **11**, 1–21 (1969).
9. Dixon, W. J. *Biometrics* **9**, 74–89 (1953).

Organelle movement in axons depends on ATP

Richard J. Adams

MRC Cell Biophysics Unit, 26–29 Drury Lane, London WC2B 5RL, UK

When living axons are examined in the light microscope many of the membrane-bounded organelles they contain can be seen to undergo rapid saltatory movements, predominantly in the retrograde direction[1]. This is thought to be an important component in the fast transport of materials in axons, the mechanism of which is unknown although cytoskeletal elements such as actin filaments and microtubules are probably involved (see ref. 2 for review and refs). I describe here experiments in which giant axons from the legs of the crab *Carcinus maenas* have been made permeable to ions and small molecules using a high voltage discharge to puncture holes in the plasma membrane[3]. This allows, in a suitable buffer, movement to be arrested due to the loss of metabolites and then reactivated by the addition of exogenous ATP, thus showing directly for the first time that rapid saltatory motion in axons is an ATP-requiring process.

Electrically active giant axons (~30 μm in diameter) from the walking legs of *C. maenas* were dissected in sea water or in a standard reactivation buffer (400 mM K glutamate, 150 mM glycine, 10 mM EGTA, 20 mM HEPES pH 7.2, 15 mM $MgCl_2$) at room temperature (about 20 °C). This was carried out in a 5-cm plastic Petri dish in the base of which a 1-cm hole had been drilled and a poly-L-lysine-coated[4] glass coverslip inserted. The polylysine allows the axons to adhere strongly to the glass coverslip so that high magnification phase contrast objectives may be used. A region of axon with particles exhibiting saltatory movement was selected for permeabilization, and the dielectric breakdown of the plasma membrane was then brought about by the discharge of a 1 μF capacitor across the axon between two fine stainless steel electrodes held in micromanipulators. A discharge voltage of 70–80 V with the electrodes set 160 μm apart (about 5 kV cm^{-1}) gave the best results and a series of holes were made in this way at intervals of 10–20 μm over a distance of 2 mm. The exact size of the holes produced is not known but can be inferred, from the results shown below, to allow molecules of molecular weight at least 900 to pass into and out of the axon; tracer studies

0028–0836/82/210327—03$01.00

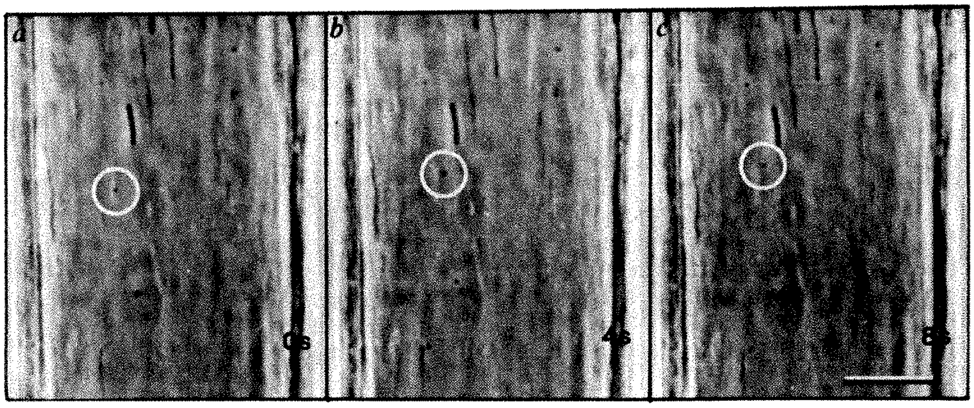
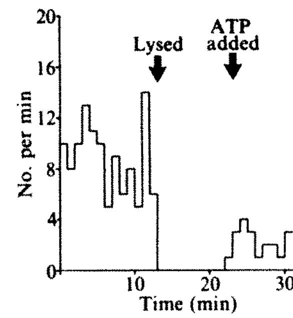


Fig. 1 A sequence of photographs of a particle (encircled) moving in an unlysed axon. The axon was mounted as described in the text and viewed with a Zeiss IM35 inverted microscope. Photographs were taken at the indicated times using Ilford FP4 film. Scale bar, 10 μm.

with adrenal medullary cells[3], have shown that this procedure produces holes of about 4 nm diameter. The stability of the holes made in the axon was shown by the observation that the axoplasm of permeabilized but not intact axons was rapidly liquified by the addition of millimolar concentrations of calcium (in the absence of protease inhibitors). It could thus be shown that the holes remain open for at least 2 h; experiments were not continued for longer than this as both permeabilized and intact axons seems to lose their ability to sustain movement after this time *in vitro*.

Movements in axons dissected from the crab leg and mounted in reactivation buffer were similar to those described in other systems[1,5]. Particles underwent transient movements at rates of up to 2 $\mu m\,s^{-1}$ with the majority (>80%) travelling in the direction of the cell body (Fig. 1). Many particles that did not undergo persistent movement underwent oscillations of small amplitude (<1 μm) along the axis of the axon. Mitochondria, which in these axons reach up to 50 μm in length, moved only occasionally, but at high rates when they did so. On permeabilization in reactivation buffer lacking ATP, movement progressively declined, with first the large and then the smaller particles coming to rest; the mitochondria often fragmented to give the appearance of beads on a string. The decline in movement depended on the loss of ATP, ADP and arginine phosphate by diffusion into the external medium. The method given above resulted in a complete cessation within about 5 min of making the holes. The decline could be accelerated, however, by the inclusion of 10 mM 2-deoxyglucose in the medium which greatly increased the rate of ATP (and thus arginine phosphate) turnover; 2-deoxyglucose was not routinely used.

Fig. 2 Number of particles passing a cross-section of axoplasm perpendicular to the long axis of the axon. The graph shows particles moving in a single axon before permeabilization ('lysis'), after permeabilization, and after reaction with 6 mM ATP.



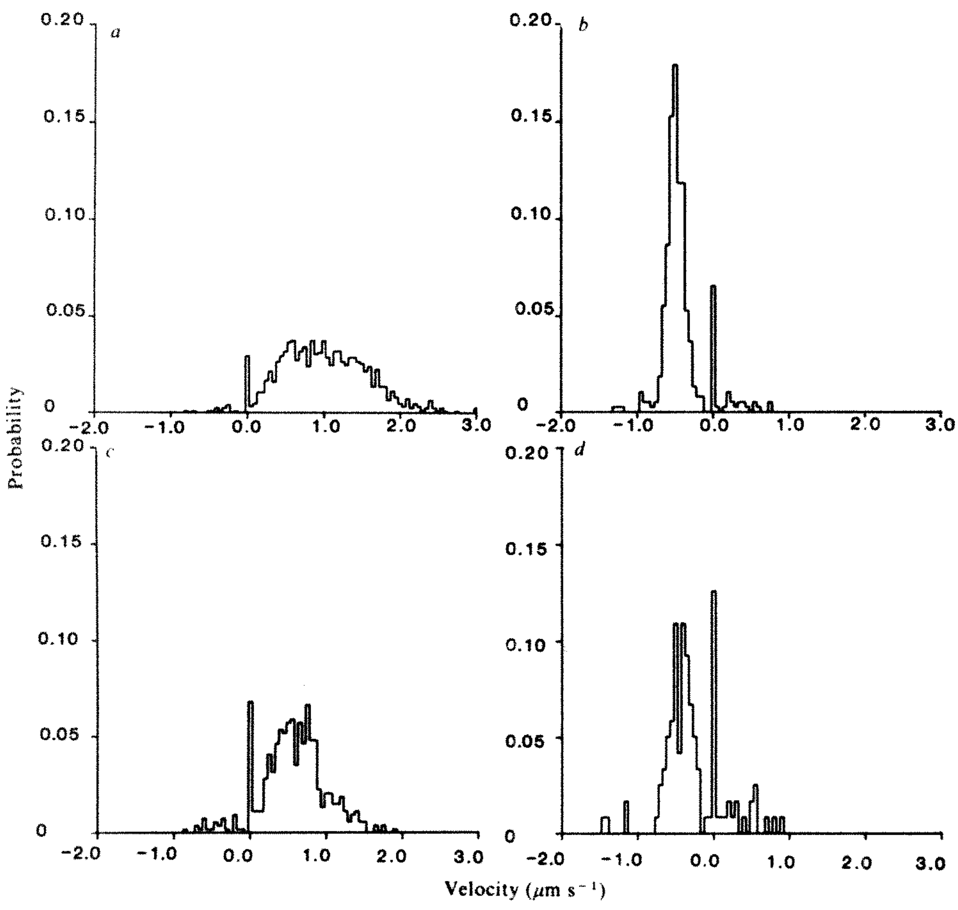


Fig. 3 Instantaneous velocities of particles moving in: *a*, *b*, intact axons; *c*, *d*, axons reactivated in 6 mM ATP. Particles are grouped according to their major direction of transport: *a*, *c*, towards the cell body (retrogradely); *b*, *d*, away, from the cell body (anterogradely). Instantaneous velocities were calculated from the time particles took to pass between gradations on an eyepiece graticule corresponding to 0.97 μm along the axon. Times were recorded initially on a cassette tape and then analysed by a computer program which calculates the velocity over each increment. The graphs represent the normalized distribution of velocities from several particles in each case: *a*, 1,251 velocity measurements from 18 particles; *b*, 380 from 5 particles; *c*, 542 from 15 particles; *d*, 119 from 5 particles.

Movement was rapidly restored by the addition of ATP; the usual concentration used was 6 mM but as little as 500 μM could restore some movement. Movement could be reactivated in axons within 90 min following permeabilization; and persisted for up to 2 h in exogenous ATP. The extent of reactivation, as measured by the number of particles passing a point per minute, varied between experiments, ranging over 25–70% (Fig. 2). This variation may be due to both the more erratic nature of movements after reactivation and a reduction in the average velocity at which particles moved. A study of instantaneous velocities (here defined as the velocity over a distance of 1 μm) of particles moving in unlysed and reactivated axons showed a reduction in the mean velocity of retrogradely moving particles (Fig. 3). The average velocity may have decreased because the conditions of reactivation were less than optimal, although ATP did not seem to have a limiting effect as further increases in its concentration did not increase the rate of movement.

Movement was not reactivated by either GTP (6 μM) or the non-hydrolysable ATP analogue adenylylimidodiphosphate (AMPPNP, 6 mM). However, the addition of 6 mM ADP did restore some movement. This may have been due to the action of adenylate kinase, a ubiquitous enzyme which can produce ATP from ADP (2 ADP$\rightleftarrows$ATP+AMP), as the restoration of movement was prevented by an inhibitor of adenylate kinase (500 μm P^1, P^5-di(adenosine-5′)pentaphosphate[6]). This inhibitor (diadenyl pentaphosphate) did not inhibit movement in the presence of exogenous ATP. It appears that the functional integrity of the axoplasm depends on the anion included in the reactivation mixture; movement continued if glutamate was replaced with aspartate but not with α-ketoglutarate, tartrate, methanesulphonate or chloride ions. However, L-glutamate could be replaced by D-glutamate without loss of activity, suggesting that a specific interaction is not involved but that merely stabilization of the structure occurs, as predicted from their relative positions in the lyotropic series. Movement in permeabilized (but not intact) axons was completely inhibited by vanadate ions at a concentration of 500 μM, although a reduction in the number of particles moving in reactivated axons was seen at concentrations as low as 5 μM. Vanadate ions in this range prevent mitosis[7] and ciliary beating[8] and inhibit the activity of a number of ATPases and kinases[9].

The buffer in which reactivation was achieved did not contain exogenously added Ca^{2+} and included 10 mM EGTA. An increase in the EGTA concentration to 50 mM had no apparent effect on reactivation. Although it remains possible that a local release of Ca^{+2} occurs from the organelles themselves and is required for them to move, this seems unlikely considering the long duration (up to 2 h) in which movement can continue in EGTA. Inclusion of the calmodulin inhibitor trifluoroperazine to a concentration of 50μM does not inhibit movement although at a concentration of 100 μM movement did appear to decline after ~15 min. It seems unlikely, due to the delay in the effect, that this is a specific inhibition but it may reflect the detergent-like activity of the drug as seen in other systems[10]. If Ca^{2+} was added to produce a concentration estimated to be ~1.5 μM (9 mM $CaCl_2$, 10 mM EGTA, 15 mM $MgCl_2$, pH 7.2) transport still continued for longer than 1 h. The addition of Ca^{2+} at concentrations higher than this is complicated by the presence of Ca^{+2}-activated proteases in the axon which cause its rapid dissolution. However, when the reactivation buffer was supplemented with protease inhibitors (1 mM TLCK, 0.4 μM acetyl-L-leucyl-L-leucyl-L-arginal, (leupeptin), 0.5 mM phenylmethyl sulphonyl fluoride) movement continued in 0.5 mM Ca^{2+}, in the absence of EGTA, for greater than 1 h. Thus this movement apparently neither requires Ca^{2+} nor is inhibited by high concentrations of Ca^{2+}.

A major aim in the study of cell motility is the establishment of cell-free systems in which movements of interest take place. Only in this way can the components required for movement be analysed and the mechanism of their action determined. The preparation described here is a first controlled step towards a cell-free system that exhibits fast axonal transport, allowing rapid saltatory movements to occur in axons that have been permeated by a defined mixture of ions and small molecules. The principal conclusion obtained from this system is that the rapid movement of organelles in axons has an absolute requirement for ATP. Furthermore, it seems likely that the hydrolysis of ATP is an essential step in the generation of movement, as it is in the contraction of muscle, the beating of cilia and cytoplasmic streaming in giant algal cells[11].

I thank Professor P. F. Baker and Dr D. Bray for much help and advice. This work was supported by an MRC studentship.

Received 22 December 1981; accepted 1 April 1982.

1. Cooper, P. D. & Smith, R. S. *J. Physiol., Lond.* **242**, 77–79 (1974).
2. Grafstein, B. & Forman, D. S. *Physiol. Rev.* **60**, 1167–1283 (1980).
3. Baker, P. F. & Knight, D. E. *Nature* **276**, 620–622 (1978).
4. Collins, F. *Devl Biol.* **65**, 50–57 (1978).
5. Smith, R. S. *Soc. Neurosci. Abstr.* **3**, 31 (1977).
6. Lienhard, G. E. & Secemski, I. I. *J. biol. Chem.* **248**, 1121–1123 (1973).
7. Cande, W. Z. & Wolniak, S. M. *J. Cell Biol.* **79**, 573–580 (1978).
8. Gibbons, I. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 2220–2224 (1978).
9. Macara, I. G. *Trends biochem. Sci.* **5**, 92–94 (1980).
10. Baker, P. F. & Knight, D. E. *Phil. Trans. R. Soc.* **B296**, 83–103 (1981).
11. Williamson, R. E. *J. Cell Sci.* **17**, 655–668 (1975).

IgE-dependent release of leukotriene C_4 from alveolar macrophages

John A. Rankin*, Margaret Hitchcock†, William Merrill*, Michael K. Bach‡, John R. Brashler‡ & Philip W. Askenase*

* Department of Medicine, Yale University School of Medicine, PO Box 3333, 333 Cedar Street, New Haven, Connecticut 06510, USA
† Department of Pharmacology, Yale University School of Medicine, The John B. Pierce Foundation, 290 Congress Avenue, New Haven, Connecticut 06510, USA
‡ Hypersensitivity Diseases Research, The Upjohn Company, Kalamazoo, Michigan 49001, USA

Slow reacting substances (SRS) are potent bronchoconstrictors that are thought to be important in the pathophysiology of asthma[1]. Human[2–5] and animal[6,7] lung fragments release SRS when sensitized by IgE antibody and challenged with specific antigen. SRS have been shown recently to be a family of peptidolipids called leukotrienes (LTC_4, LTD_4 and LTE_4) that are derived from arachidonic acid[8–11] and are potent bronchoconstrictors *in vivo* and *in vitro*[12–14]. It is not known which cell(s) is responsible for releasing SRS but two recent observations have directed our attention to alveolar macrophages. First, Capron and co-workers demonstrated that serum IgE antibody from rats immune to *Schistosoma mansoni* parasites could cause rat peritoneal macrophages to be cytotoxic for schistosomules *in vitro*[15]. This is a result of the interaction of IgE with macrophage IgE Fc receptors[16]. Similar receptors are also present on rat alveolar macrophages[17]. Second, Bach and Brashler[18–21] established that rat peritoneal macrophages could release a mixture of LTC_4 and LTD_4 when stimulated by the calcium ionophore A23187. We now show that rat alveolar macrophages can be activated by calcium ionophore to release SRS and, when challenged with IgE antibody and specific antigen, release LTC_4.

Male Sprague–Dawley rats were killed in a CO_2-filled chamber. Their lungs were lavaged with 150 ml of 0.9% NaCl and the recovered cell suspensions were centrifuged at 500g for 8 min. The cell pellet was washed with 0.9% NaCl, centrifuged and resuspended in Hank's balanced salt solution (HBSS); in mM: $CaCl_2$ 0.95, KCl 5.3, KH_2PO_4 0.4, $MgCl_2$ 149, $MgSO_4$ 0.04, NaCl 136.8, $NaHCO_3$ 4.1, Na_2HPO_4 0.33, glucose 55). The macrophages were identified by uptake of neutral red dye and total cells were counted in a haemocytometer. Viability was assessed by Trypan blue exclusion and the cell number was adjusted to 5×10^6 viable macrophages per ml in HBSS.

0028–0836/82/210329—03$01.00

Cell differentials on Wright–Giemsa-stained cytocentrifuged preparations revealed the lung lavages to contain 93.2% macrophages, 3.3% lymphocytes, 3.3% neutrophils and 0.14% mast-basophiloid cells (data represent the mean of 15 experiments).

The ability of rat alveolar macrophages to release SRS was tested initially by incubating cell suspensions for 20 min with 1 μmol of the calcium ionophore A23187, in the presence of 5×10^{-4} M L-cysteine. In subsequent experiments, cell suspensions were stimulated with purified mouse monoclonal anti-DNP (dinitrophenyl) IgE antibody[22] (provided by Dr D. H. Katz) and DNP–human serum albumin ($DNP_{22}HSA$) in the following manner: cell suspensions were incubated with anti-DNP IgE antibody for 20 min at 37 °C, $DNP_{22}HSA$ was then added and incubation continued for a further 20 min.

Controls consisting of cells alone and cells with antibody alone were incubated simultaneously. Following incubation, the cell suspensions were centrifuged at 500*g* and the supernatants assayed for SRS on guinea pig ileum in the presence of atropine sulphate (3.4 $\mu g\ ml^{-1}$), methysergide maleate (0.1 $\mu g\ ml^{-1}$) and mepyramine maleate (0.34 $\mu g\ ml^{-1}$). Positive contractions were repeated in the presence of the selective SRS antagonist[23] FPL 55712 (0.1 $\mu g\ ml^{-1}$). As the chemical nature of the SRS released by alveolar macrophages was unknown initially, we compared quantitatively the magnitude of the contractions produced by our supernatants with those produced by synthetic (5*S*, 6*R*, 7E, 11*Z*, 14*Z*)-5-hydroxy-6-[(2*R*-amino-2-carboxyethyl)thio]7,9,11,14-eicosatetraenoic acid[24] (LTE_4; kindly made available by Dr M. Rosenberger, Hoffman-LaRoche). LTE_4 is a constituent of SRS that appears with biologically generated LTC_4 and LTD_4.[25] Thus, our data are expressed in $mol\times10^{-10}$ of LTE_4 equivalents per 5×10^6 alveolar macrophages.

Figure 1 shows that the non-immunological stimulus A23187 induced the release of SRS ($3\pm1.1\times10^{-10}$ mol of LTE_4 equivalents). Cells incubated in identical conditions without A23187 and L-cysteine showed no spontaneous release of SRS. Thus, rat alveolar macrophages can release SRS when stimulated by the calcium ionophore.

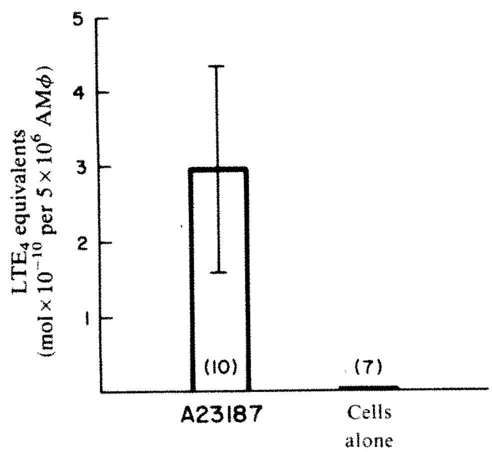


Fig. 1 A23187-induced release of SRS from alveolar macrophages. Rat lung lavages consisting of 93.2% alveolar macrophages and only 0.14% mast-basophiloid cells were adjusted to 5×10^6 viable alveolar macrophages per ml and were incubated at 37 °C with 5×10^{-4} M L-cysteine for 150 s and then for 18 min with 1 μM A23187. After centrifugation the supernatants were assayed for SRS on guinea pig ileum in the presence of atropine (3.4 $\mu g\ ml^{-1}$), mepyramine (0.34 $\mu g\ ml^{-1}$) and methysergide (0.1 $\mu g\ ml^{-1}$). Positive contractions were repeated in the presence of 0.1 $\mu g\ ml^{-1}$ FPL 55712 and were inhibited by 80–100%. Measured contractions were compared with a synthetic LTE_4 standard. LTE_4 equivalents are shown on the ordinate in $mol\times10^{-10}$ per 5×10^6 alveolar macrophages. The columns represent mean values for the experiments performed. The number of experiments is shown at the base of the columns; bars represent standard errors.

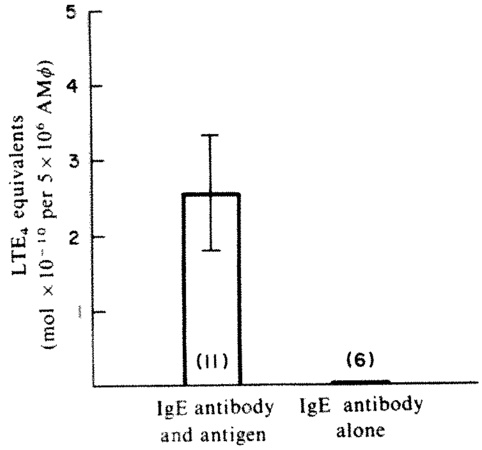


Fig. 2 SRS released from rat alveolar macrophages: induction by mouse monoclonal anti-DNP IgE and $DNP_{22}HSA$. Cells were prepared and SRS assayed as described in Fig. 1 legend. Positive contractions were inhibited by 100% in the presence of 0.1 $\mu g\ ml^{-1}$ FPL 55712. LTE_4 equivalents are shown on the ordinate in $mol\times10^{-10}$ per 5×10^6 alveolar macrophages. The columns represent the mean values for SRS released, and the bars the standard error. The number of experiments is shown at base of columns. Alveolar macrophages were incubated with hapten affinity-purified anti-DNP IgE antibody (10 μg) for 20 min at 37 °C and for an additional 20 min with $DNP_{22}HSA$ (100 ng). In control experiments, alveolar macrophages were incubated with IgE alone.

Figure 2 shows the pooled results of 11 experiments in which cell suspensions were incubated with anti-DNP IgE for 20 min, followed by a further 20 min incubation with $DNP_{22}HSA$. Alveolar macrophages also released SRS by this IgE-dependent reaction ($2.5\pm0.8\times10^{-10}$ mol of LTE_4 equivalents). When cells were incubated with antibody alone (Fig. 2), or with antigen alone, no SRS was released (data not shown). Thus, immunological release of SRS by alveolar macrophages probably depends on a combination of specific IgE antibody and antigen.

Supernatants generated in additional experiments using anti-DNP IgE and $DNP_{22}HSA$ were subjected to chemical analysis in order to characterize the nature of the SRS. Pooled supernatants were deproteinized by 1:4 dilution with methanol and centrifugation, stored at −70 °C, thawed, concentrated, hydrolysed in 0.1 M NaOH, desalted and then chromatographed on a Florisil column with elution in a methanol/water gradient[19]. All biological activity eluted in the same position as synthetic LTC_4 standard (prepared by Dr D. R. Morton, Upjohn) and LTC_4 derived from rat peritoneal macrophages[19,21]. The single peak was concentrated and applied with ^3H-labelled LTC_4 tracer (NEN) to a reverse-phase HPLC column (10 μm Nucleosil C-18), eluted isocratically with 65% methanol, 35% water and 0.01% acetic acid, *p*H 5.4. The absorption at 280 nm, the biological activity and the radioactivity all co-eluted from the column, while a synthetic LTD_4 standard was clearly resolved from the elution peak of the radiolabelled LTC_4 in a separate run on the same column. Thus, two different chromatographic procedures showed that supernatants from alveolar macrophages, activated by an IgE-dependent reaction, contained predominantly LTC_4.

We have thus demonstrated that (1) rat alveolar macrophages release SRS when stimulated non-specifically by the calcium ionophore A23187 in the presence of L-cysteine, and (2) IgE antibody and appropriate antigen cause alveolar macrophages to release SRS leukotriene, LTC_4. Calcium ionophore has been used previously to elicit SRS release from mouse and rat peritoneal mononuclear cells[18,26], human leukocytes[27], rat basophilic leukaemia cells[28,29], and mouse mastocytoma cells[30]. However, our results differ from previous studies in two ways. First, Orange *et al.*[26], using a different strain of rat and a different concentration of L-cysteine, elicited "very little" SRS from alveolar macrophages with A23187. We calculate that the total amount produced by IgE-stimulated alveolar macrophages is

of the same order of magnitude as produced by ionophore-stimulated rat peritoneal macrophages. Second, Bach *et al.*, in collaboration with Capron and Joseph, reported that activation of rat peritoneal macrophages by IgE–anti-IgE—conditions that cause release of lysosomal enzymes[31]—failed to elicit SRS release[32]. We have, in fact, confirmed this result: when rat peritoneal cells were incubated with anti-DNP IgE and $DNP_{22}HSA$ in conditions identical to those used for our alveolar macrophages, they did not release SRS (data not shown). Thus, alveolar macrophages may be specialized to perform this function. Furthermore, alveolar macrophages release predominantly LTC_4 in an IgE-dependent mechanism whereas rat peritoneal macrophages when stimulated with A23187, release both LTC_4 and LTD_4[20,21]. In our experiments, it is probable that alveolar macrophages are the source of SRS because alveolar macrophages comprised 93% of the cells in the lung lavage, and when lavage cells were first allowed to adhere to glass for 1 h they were also capable of IgE-dependent release of SRS (data not shown). However, we cannot exclude the possibility that the minority cells present in our cell suspensions (such as mast-basophiloid cells) contributed to the detected SRS either directly[33] or indirectly[1].

The exact mechanism by which IgE antibody and antigen effect SRS release remains to be determined. Capron and co-workers established that complexes of IgE and schistosome antigen were responsible for eliciting cytotoxicity for schistosome larvae in rat periotoneal macrophages[34] and Dessaint and co-workers showed that IgE binding to Fc receptors on these cells is of much lower affinity than IgE binding to mast cells[16]. Thus IgE antibody and antigen may effect SRS release from alveolar macrophages through formation of immune complexes that activate the cells via these lower-affinity Fc receptors. However, it is also possible that a cytophilic mechanism, analogous to that which sensitizes mast cells and basophils for IgE-dependent release of histamine, is responsible, as 'direct' cytophilic antibodies with high affinity for murine macrophage Fc receptors do exist—but they may be specific for IgG2a[35]. Our preliminary experiments indicate that both mechanisms may be involved because washing the cells after incubation with IgE and before addition of antigen reduces, but does not eliminate, SRS release. Further experiments are being conducted to investigate this issue further.

The relevance of our findings to allergic lung disease mechanisms remains to be determined. It is clear that many important mediators are released by mast cell-dependent pathways[1] and there is recent evidence that pulmonary mast cells may be an important source of SRS in humans[33]. However, it has also been demonstrated recently that human alveolar macrophages can release lysosomal enzymes and superoxide anion after incubation with IgE and anti-IgE antibody, or with serum from allergic patients and the appropriate allergen[36]. Merrill and co-workers have examined the immunoglobulins present in the lung lining fluid of the lower respiratory tracts of normal individuals, and found that although IgG and IgA were the predominant immunoglobulins present, IgE was present in detectable quantities in ~80% of the 43 people tested[37]. Similar studies on asthmatic individuals have not been reported, but IgE levels are increased in the nasal secretions of patients with allergic rhinitis[38–40] and probably result from local production[41]. As the mucosa of the nose and lower respiratory passages are in many ways similar, it is reasonable to hypothesize that IgE–antigen–macrophage interactions occur on the air–surface interface of the lung. Our demonstration that rat alveolar macrophages release SRS by an IgE-dependent mechanism raises the possibility that IgE-dependent release of mediators by alveolar macrophages may have a role in asthma or other immunologically mediated lung diseases.

This work was supported in part by NIH grants AI-12211, AI-11077, AI-10497, AI-17555 and 5T 32-Es-07086, the Council for Tobacco Research and Hoffman La-Roche Inc. J.A.R. is a William Wirt Winchester fellow in the pulmonary section.

Received 5 June 1981; accepted 26 March 1982.

1. Lewis, R. A. & Austen, K. F. *Nature* **293**, 103–108 (1981).
2. Parish, W. E. *Nature* **215**, 738–739 (1967).
3. Sheard, P., Killingback, P. G. & Blair, A. M. J. N. *Nature* **216**, 283–284 (1967).
4. Orange, R. P., Austen, W. G. & Austen, K. F. *J. exp. Med.* **134**, 136S (1971).
5. Kay, A. B. & Austen, K. F. *J. Immun.* **107**, 899–902 (1971).
6. Goodfriend, L., Kovacs, B. A. & Rose, B. *Int. Archs Allergy appl. Immun.* **30**, 511–518 (1966).
7. Ishizaka, T., Ishizaka, K., Orange, R. P. & Austen, K. F. *J. Immun.* **104**, 335–343 (1970).
8. Hammarström, S. *et al. Biochem. biophys. Res. Commun.* **91**, 1266–1272 (1979).
9. Morris, H. R., Taylor, G. W., Piper, P. J. & Tippins, J. R. *Nature* **285**, 104–106 (1980).
10. Watanabe-Kohno, S. & Parker, C. S. *J. Immun.* **125**, 946–955 (1980).
11. Goetzel, E. J. *New Engl. J. Med.* **14**, 822–825 (1980).
12. Drazen, J. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 4354–4358 (1980).
13. Dahlen, S.E., Hedqvist, P., Hammarström, S. & Samuelsson, B. *Nature* **288**, 484–486 (1980).
14. Hanna, C. J., Bach, M. K., Pare, P. D. & Schellenberg, R. R. *Nature* **290**, 343–344 (1981).
15. Capron, A., Dessaint, J.-P., Capron, M. & Bazin, H. *Nature* **253**, 474–475 (1975).
16. Dessaint, J.-P. Torpier, G., Capron, M., Bazin, H. & Capron, A. *Cell. Immun.* **46**, 12–13 (1979).
17. Boltz-Nitulescu, G., Bazin, H. & Spiegelberg, H. L. *J. exp. Med.* **154**, 373–384 (1981).
18. Bach, M. & Brashler, J. R. *J. Immun.* **120**, 998–1005 (1978).
19. Bach, M. K., Brashler, J. R., Brooks, C. D. & Neerken, A. J. *J. Immun.* **122**, 160–165 (1979).
20. Bach, M. K., Brashler, J. R., Hammarström, S. & Samuelsson, B. *Biochem. biophys. Res. Commun.* **93**, 1121–1126 (1980).
21. Bach, M. K., Brashler, J. R., Hammarström, S. & Samuelsson, B. *J. Immun.* **125**, 115–117 (1980).
22. Liu, F. T. *et al. J. Immun.* **124**, 2728–2737 (1980).
23. Augstein, J., Farmer, J. B., Lee, T. B., Sheard, P. & Tattersoll, M. L. *Nature new Biol.* **45**, 215–217 (1973).
24. Rosenberger, M. & Neukom, C. J. *J. Am. chem. Soc.* **102**, 5426 (1980).
25. Lewis, R. A., Drazen, J. M., Austen, F. K., Clark, D. A. & Corey, E. J. *Biochem. biophys. Res. Commun.* **96**, 271–277 (1980).
26. Orange, R. P., Moore, E. G. & Gelfand, F. W. *J. Immun.* **124**, 2264–2267 (1980).
27. Conroy, M. C., Orange, R. P. & Lichtenstein, L. M. *J. Immun.* **116**, 1677–1681 (1975).
28. Jakschik, B. A., Kulczycki, A. Jr, MacDonald, H. H. & Parker, C. W. *J. Immun.* **119**, 618–622 (1977).
29. Morris, H. R. & Taylor, G. W. *Prostaglandins* **19**, 185–201 (1980).
30. Murphy, R. C., Hammarström, S. & Samuelsson, B. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4275–4279 (1979).
31. Dessaint, J.-P., Capron, A., Joseph, M. & Bazin, H. *Cell. Immun.* **46**, 24–34 (1979).
32. Bach, M. K., Brashler, J. R., Johnson, M. A. & Drazen, J. M. in *Biochemistry of Allergic Reactions*, 37–50 (Liss, New York, 1981).
33. MacGlashan, D. W. Jr *et al. Fedn Proc.* **41**, 375 (1982).
34. Capron, A., Dessaint, J.-P., Rousseau, R., Capron, M. & Bazin, H. *Eur. J. Immun.* **7**, 315–322 (1977).
35. Unkeless, J. C. & Eisen, H. *J. exp. Med.* **142**, 1520–1533 (1975).
36. Joseph, M., Tonnel, A. B., Capron, A. & Voison, C. *Clin. exp. Immun.* **40**, 416–422 (1980).
37. Merrill, W. W., Naegel, G. P. & Reynolds, H. Y. *J. lab. clin. Med.* **96**, 494–500 (1980).
38. Hobday, J. A., Cahe, M. & Turner, K. J. *Clin. exp. Immun.* **9**, 577–583 (1971).
39. Nakejima, S., Gillespie, D. N. & Gleich, G. J. *Clin. exp. Immun.* **21**, 306–317 (1976).
40. Yunginer, J. W. & Gleich, G. J. *J. Allergy clin. Immun.* **51**, 174–186 (1973).
41. Platts-Mills, T. A. E. *J. Immun.* **122**, 2218–2225 (1979).

Corticotropin releasing factor produces behavioural activation in rats

Richard E. Sutton, George F. Koob*, Michel Le Moal*, Jean Rivier & Wylie Vale

Peptide Biology Laboratory and * Arthur Vining Davis Center for Behavioral Neurobiology, The Salk Institute, PO Box 85800, San Diego, California 92138, USA

A 41-residue peptide with high potency and intrinsic activity to stimulate the secretion of corticotropin and β-endorphin by the adenohypophysis has been characterized and synthesized[1,2,19]. From experience with other hypophysiotropic and brain peptides[3–5], and studies suggesting that CRF could act centrally to activate the sympathetic nervous system (M. Brown *et al.*, in preparation), we considered it likely that this corticotropin/β-endorphin releasing factor (CRF) would have direct behavioural actions. We decided to investigate the effects of CRF on a simple measure of activation in the rat, locomotor activity in photocell cages. We report here that centrally injected synthetic CRF produces a dose-dependent locomotor activation in rats. This increase in activity was not produced by peripheral administration of CRF. In an open field test, rats receiving centrally injected CRF exhibited behaviour consistent with an increase in emotionality. These results suggest that CRF may have an activating action in the central nervous system independent of its effects on the anterior pituitary gland.

0028-0836/82/210331—03$01.00

Table 1 Effects of CRF (i.c.v.) on behaviour of rats in open field

	Crossings* inner	Crossings outer	Rearing* free	Rearing wall	Rearing duration (s)	Grooming duration
Saline	7.9±3.0	96.4±7.6	4.9±1.4	33.3±3.7	44.0±4.7	29.4±6.6
CRF (0.0015 nmol)	14.6±4.6†	90.1±10.4	9.4±3.3	34.9±3.7	42.1±4.1	38.1±14.2
CRF (0.015 nmol)	5.1±1.3	92.3±13.4	2.3±0.6	24.4±3.6	37.8±8.1	30.6±8.9
CRF (0.15 nmol)	1.9±0.6	29.1±7.5‡	0.0±0.0‡	4.3±1.5‡	7.1±2.3‡	90.9±18.4‡

* F_{max} $P<0.05$: raw data square root transformed for ANOVA; *a posteriori* tests were only made after obtaining significant differences using ANOVA. † Significantly different from CRF 0.15 nmol group, Newman–Keuls test $P<0.05$. ‡Significantly different from saline group, Newman–Keuls test $P<0.05$.

In the initial experiments the subjects were male albino Wistar rats (200–230 g at the time of injection) which were group housed and maintained in a temperature- and light-controlled environment. For the intracerebroventricular (i.c.v.) injections, rats were equipped with a cannula aimed above the lateral ventricle. For this surgery, the animals were anaesthetized with Chloropent (Fort Dodge Laboratories) and secured in a Kopf stereotaxic instrument with guinea pig ear bars. A 7-mm long, 23-gauge stainless steel guide cannula was lowered within 1 mm of the ventricle and anchored in place with two stainless steel screws and dental cement. With the tooth bar 5 mm above interaural zero, coordinates were −0.6 mm posterior to bregma, ±2.0 mm lateral and −3.2 mm below the skull surface at the point of entry. For an injection, the dummy stylet was removed and an 8-mm 30-gauge stainless steel cannula with 1 m of Tygon microbore tubing (Norton Plastics and Synthetics) was inserted through the guide.

CRF, synthesized by the solid phase method on a paramethylbenzylhydryl amine resin using previously described techniques[6], was purified by HPLC after cleavage and deprotection by HF. The isolated material was shown to be highly purified (>98%) when tested by homogeneity in several reverse phase HPLC systems including analysis of a tryptic object. As reported earlier, such closely related analogues of CRF as des-Ala^{41}-CRF, [$Met(O)^{21}$]-CRF, des-$Ileu^{40}$, Ala^{41}-CRF and CRF-OH could be separated in the chromatographic conditions developed in our laboratory (ref.1 and J.R. *et al.*, in preparation). One hour before injections, the peptide was dissolved in 0.9% saline and kept on ice; 2 µl of peptide were injected by gravity over a 30-s period simply by raising the tubing above the head of the rat until flow began. Volume was measured by marks on the PEID tubing previously calibrated with a 10-µl Hamilton syringe. Only those rats whose cannulae flowed easily with this technique were used in the experiment. The lateral ventricle cannula placements were verified by injecting 2 µl of toludine blue dye into the ventricle and immediately killing the rat by decapitation. Of 35 rats tested in this way, all showed spread of the dye throughout the ventricular system. For subcutaneous CRF injections, each rat received the peptide dissolved in 1 ml per kg body weight saline in the neck.

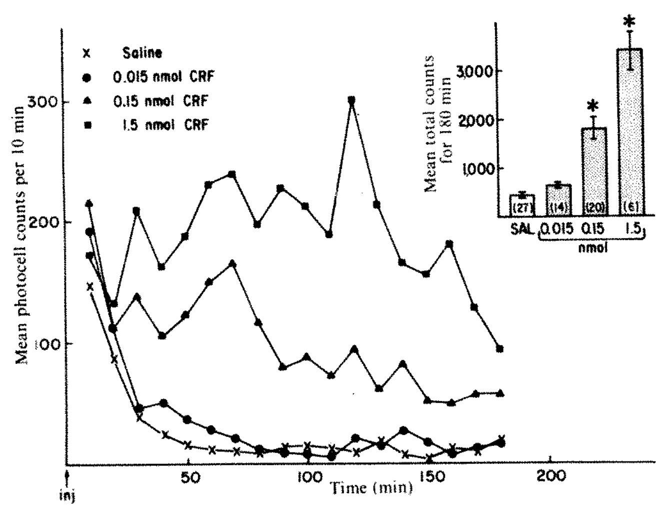


Fig. 1 Locomotor response after i.c.v. infusions of 0.1, 1 and 10 µg corresponding to 0.015, 0.15 and 1.5 nmol of CRF, respectively. Ordinate refers to total photocell counts for each 10-min period of a 3-h test.* Significantly different from saline, and from each other. $P<0.05$, Neuman-Keuls test following analysis of variance. Mean total counts for 180 min represent average ±s.e.m.

Locomotor activity was measured in a 4×4 bank of 16 wire cages 20×25×36 cm each with two horizontal IR photocell beams across the long axis 2 cm above the floor. Noncumulative photocell beam interruptions were recorded every 10 min from electromechanical counters located in an adjacent room. Each rat was habituated first overnight and then for 2 h 1 day before peptide injections. On a test day, rats were habituated for 90 min just before peptide injections, which were given at 10 a.m. Results were analysed with a two-factor analysis of variance (ANOVA), with peptide or saline groups being the independent factor and time the repeated measure. If ANOVA revealed a significant ($P<0.05$) group effect, comparisons of individual means were made using the Newman–Keuls *a posteriori* test.

Open field testing took place in a large open field[7] measuring 115×115×46 cm tall and divided by thin white lines into 23×23 cm squares. In the open field, the light intensity was ~150 ft-candles and the white noise level was 70 dB. The open field was enclosed by sheets of one-way viewing sunscreen. Naive rats were tested in a 5-min session by placing the rat in the centre of the open field. The following behavioural measures were taken by two blind observers: rearing—inner and outer squares; locomotion—inner and outer squares, time rearing, time grooming and number of grooming bouts. Here, the rats were injected with 0.0015, 0.015 and 0.15 nmol of CRF 1 h before testing.

CRF injected i.c.v. produced a dose-dependent increase in activity as measured by photocell interruptions (Fig. 1). This increase in activity was significant at the 0.15 and 1.5 nmol doses. Rats injected with saline habituated rapidly, showing low photocell counts after 1 h, but rats given 1.5 nmol CRF had high counts for over 5 h post-injection (data not shown). Repeated i.c.v injections of CRF over several days did not modify this increase in behavioural activity. Following injections, rats in the photocell cages were observed through a one-way mirror. In addition to normal locomotion and sniffing, the rats injected with 0.15 and 1.5 nmol CRF exhibited a unique set of behaviours which seemed to reflect a general behavioural activation. These included elevated walking, grooming and rearing. The elevated walking was characterized by rhythmically walking forwards and backwards such that only the toes of the rat touched the cage. The rats frequently climbed up and down the two mesh sides of the photocell cages, and they also engaged in other repetitive locomotion such as moving forwards and backwards in a straight line with their head along the ground and sniffing vigorously. While rearing, rats pawed rapidly against the sides of the cages. When returned to their home cages, the rats exhibited an increase in mounting behaviour and aggressivity.

Although other peptides, such as endorphins and ACTH, have been shown to produce increases in behavioural activity[8–11], the time course and nature of the response observed with CRF differed dramatically from the responses observed with endorphins. Opioid peptide activation following i.c.v. injection consists of an initial depressant phase which varies in duration with the dose and of a staring behaviour that

approaches a catatonic-like state at higher doses. CRF did not produce these responses, nor the bursts of locomotor activity interrupted by scratching and grooming behaviour observed with opioid peptides[8], nor the well-documented 'stretching–yawning syndrome' (SYS) observed after i.c.v. injection of ACTH and ACTH-like peptides[9–12]. In contrast, the activation produced by CRF at the lower doses appeared to be an exaggeration of normal activation produced by introduction into a familiar environment. At the highest dose, the behaviour was characterized by the more bizarre elevated walking and rhythmic walking forwards and backwards.

Results in the open field test indicate that rats receiving CRF i.c.v. in a novel environment respond in such a way as to be consistent with an increased emotionality or increased sensitivity to the stressful aspects of the situation. Here, the rats at the 0.15 nmol dose of CRF showed decreases in locomotion and rearing, but increases in freezing behaviour. These behavioural responses are characteristic of an increase in emotional reactivity (Table 1). Typically, a rat injected with 0.15 nmol CRF and placed 1 h later in the open field moved hesitantly to the outer squares of the open field. The animal then either slowly circled the open field, remaining close to the floor and wall with virtually no rearing, or remained in one of the corners grooming or moving forwards and backwards. In contrast, control animals injected with saline rapidly circled the open field initially with a substantial amount of rearing and later in the 5-min period made forays into the inner squares with some free rearing. Interestingly, at the lowest dose of CRF, 0.0015 nmol, there were some increases in rearing and locomotion (Table 1) that would be considered more consistent with the activation observed in the photocell cages described above.

Subcutaneous injections of identical doses of CRF failed to produce significant effects in the open field test, although the highest dose produced some trends in the same direction. These trends were not statistically significant by ANOVA, nor are these results of similar magnitude to those observed by central injections. For example, the largest effect of the peripheral injection was a 46% decrease in wall rearings in the open field at the 0.15 nmol per rat dose compared with an 87% decrease following central injection (ANOVA $P > 0.10$). Also, there was a complete blockade of the free rearing with the 0.15 nmol dose injected i.c.v. but only a 37%, not statistically significant, change following the same dose injected subcutaneously. Thus, these preliminary results suggest a relatively minor contribution of the systemic actions of the peptide for the open field effects. In addition, subcutaneous injections of 0.015, 0.15, 1.5 and 7.5 nmol of CRF failed to increase locomotor activity significantly in the photocell cages (data not shown). These results together suggest that CRF is acting centrally to produce its behavioural activation, as these systemic doses are capable of producing significant elevations of plasma ACTH and β-endorphin.

Although these results suggest that CRF can act centrally to produce a dose-dependent behavioural activation, physiologically CRF might act in concert with ACTH, β-endorphin or glucocorticoids to bring about some behavioural responses to circumstances such as stress. CRF, however, is structurally specific in its behavioural effects, as the C-terminal free carboxyl analogue, which is inactive *in vitro*[1], did not augment locomotor activity to the extent observed with CRF itself. Based on maximal response, this analogue was $\frac{1}{1,000}$th as potent as CRF in increasing locomotor activity. Sauvagine, a non-mammalian polypeptide with striking homology with CRF[13] and similar *in vitro* activity[20], was equipotent in increasing locomotor activity.

Thus, in a familiar environment (photocell cages) CRF produces a behavioural activation as reflected by a prolonged increase in locomotor activity where similarly treated, saline-injected animals rapidly habituate and go to sleep. However, in a novel stressful environment (open field test), similar doses produce decreases in behaviour such as rearing and inner square crossings that are consistent with an increase in the stress of the situation. How these effects described above extend to other less stressful situations involving such theoretical constructs as learning and attention remains to be determined.

The present results thus provide evidence that CRF may have neurotropic action independent of the hypophyseal system. These observations, in conjunction with the well documented involvement of β-endorphin and corticotropin in stress[14–18], suggest a possible role of central CRF in similar adaptive mechanisms.

The research was supported in part by NIH grants AM26741 and AA03504, and conducted in part by the Clayton Foundation for Research, California Division. W.V. is a Clayton Foundation Investigator. We thank Drs F. Bloom and M. Brown for helpful discussions, Sylvia Quijada, Ron Kaiser, Rick McClintock, R. Galyean, Eva Alwerud and Jonathan Schleimer for technical assistance, Bob Peterfreund for coding the blind procedure, and Susan McCall for manuscript preparation.

Received 5 January; accepted 29 March, 1982.

1. Vale, W., Spiess, J., Rivier, C. & Rivier, J. *Science* **213**, 1394–1397 (1981).
2. Speiss, J., Rivier, J., Rivier, C. & Vale, W. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6517–6521 (1981).
3. Witter, A. & deWied, D. in *Handbook of the Hypothalamus* (eds Morgane, P. & Panksepp, J.) 307–451 (Dekker, New York, 1980).
4. Vale, W., Rivier, C. & Brown, M. in *Handbook of the Hypothalamus* (eds Morgane, P. & Panksepp, J.) 165–252 (Dekker, New York, 1980).
5. Plotnikoff, N. P., Prange, A. J. Jr, Breese, G. R., Anderson, M. S. & Wilson, I. C. *Science* **178**, 417–418 (1972).
6. Marki, W., Spiess, J., Tache, Y., Brown, M. & Rivier, J. E. *J. Am. chem. Soc.* **103**, 3178–3185 (1981).
7. Whimbey, A. E. & Denenberg, V. H. *J. comp. Physiol. Psychol.* **63**, 500–504 (1967).
8. Segal, D. S. *et al.* in *Endorphins in Mental Health Research* (eds Usdin, E., Bunney, W. E. Jr & Kline, N. S.) 307–324 (Macmillan, London, 1979).
9. Dunn, A. J., Green, E. J. & Isaacson, R. L. *Science* **203**, 281–283 (1979).
10. Gispen, W. H., Weigant, V. M., Greven, H. M. & DeWeid, D. *Life Sci.* **17**, 645–652 (1975).
11. Ferrari, W., Gessa, G. L. & Vargiu, L. *Ann. N.Y. Acad. Sci.* **104**, 330–345 (1963).
12. DeWied, D. *Am. J. Physiol.* **207**, 255–259 (1964).
13. Montecucchi, P. C., Henschen, A. & Erspamer, V. *Hoppe-Seyler's Z. physiol. Chem.* **360**, 1178–1183 (1980).
14. Denenberg, V. H. *N.Y. Acad. Sci.* **159**, 852–859 (1969).
15. Selye, H. *Stress* (Acta, Montreal, 1950).
16. Rossier, J. *et al. Nature* **270**, 618–620 (1977).
17. Guillemin, R. *et al. Science* **197**, 1367–1369 (1977).
18. Akil, H., Watson, S. J., Barchas, J. D. & Li, C. H. *Life Sci.* **24**, 1659–1665 (1979).
19. Brown, M. *et al. Life Sci.* **30**, 207–210 (1982).
20. Brown, M. *et al. Regulatory Peptides* (in the press).

Dimeric tetrapeptide enkephalins display extraordinary selectivity for the δ opiate receptor

Yasuyuki Shimohigashi, Tommaso Costa, Hao-Chia Chen & David Rodbard

Endocrinology and Reproduction Research Branch,
National Institute of Child Health and Human Development,
National Institutes of Health, Clinical Center Room 10B17,
Bethesda, Maryland 20205, USA

The μ and δ opiate receptors, postulated on the basis of pharmacological observations, have been well characterized in ligand-binding studies[1–6]. The several classes of opiate receptor appear to subserve different functions[7–9]. Although highly specific μ ligands are available, there has been a paucity of high-affinity ligands with δ specificity that could be used to investigate the functions, properties and spatial distribution of the δ receptor. Previous structure–activity studies of the enkephalins have suggested that a free carboxyl group at the C-terminus is an important determinant of δ selectivity[9,10]: amidation of Leu⁵ or Met⁵ results in a non-selective ligand. Recently, however, we have demonstrated that cross-linking enkephalin amides by a methylene bridge of suitable length produces a dimeric pentapeptide, with greater affinity and selectivity for the δ receptor than is found in the original δ ligand[11,12]. Evidence that this dimer may interact simultaneously with two δ, but not two μ receptors[12] is consistent with independent demonstrations that the δ receptors are clustered in the membrane[7,13]. When the C-terminal amino acid of enkephalin is removed, the resulting tetrapeptide enkephalin amide H-Tyr-D-Ala-Gly-Phe-NH_2 and its analogues are potent and selective

ligands for μ opiate sites[14,15]. If the δ receptors were closely clustered, as suggested by our findings with dimeric enkephalin pentapeptides[11,12] and by others[7,13], then it might be possible to form a dimeric analogue of the tetrapeptide enkephalin which could interact with two δ receptors but not with two μ receptors. Theoretically, this would confer δ selectivity on the dimer of a μ-selective ligand. We have thus synthesized a series of dimeric analogues of the μ-selective tetrapeptide [D-Ala^2, des-Leu^5]enkephalin amide by cross-linking at the C-terminus with NH_2-$(CH_2)_n$-NH_2, with $n = 2$–12. We report here that the dimer with $n = 12$ has a nearly 1,000-fold increase in δ/μ selectivity ratio, and is thus a δ-selective ligand.

A series of dimeric tetrapeptide enkephalins have been synthesized by a two-step coupling procedure: cross-linking of Boc-Phe-OH with NH_2-$(CH_2)_n$-NH_2 (where $n = 2$–12), followed by elongation with Boc-Tyr-D-Ala-Gly-OH. Purity was confirmed by mass spectrometry, amino acid analysis and TLC. [D-Ala^2, D-Leu^5]enkephalin (DADLE), [D-Ser^2, Leu^5]enkephalyl-Thr (DSLET) and monomeric tetrapeptide [D-Ala^2, des-Leu^5]enkephalin amide (DAPEA) were purchased from Peninsula Laboratories (San Carlos, California). All peptides were evaluated for their activity in two different radioligand assay systems, using either ^{3}H-DADLE (40 Ci mmol^{-1}; 0.10 nM) and membranes from neuroblastoma–glioma hybrid (NG108-15) cells[5,16], or ^{3}H-naloxone (^{3}H-NAL, 50.2 Ci mmol^{-1}; 0.15 nM) in rat brain membrane preparations[17].

In the 'δ receptor' assay, the removal of the C-terminal amino acid residue from the pentapeptides [D-Ala^2, Leu^5]enkephalin amide (DALEA) and DADLE, giving the tetrapeptide [D-Ala^2, des-Leu^5]enkephalin amide (DAPEA), caused a substantial loss of δ activity (Table 1). When ^{3}H-DADLE was used as the labelled ligand in the assay, DAPEA was approximately 30-fold less potent than DALEA and DADLE. In contrast, the dimeric tetrapeptide enkephalins (DTE$_n$) were very active for the δ receptors with IC_{50} values of 1–4 nM. In particular, DTE$_{10}$ and DTE$_{12}$ (with cross-linking methylene chains of $n = 10$ and 12 respectively) were 30-fold more potent than the DAPEA monomer, and were as active as DADLE and DALEA, and more active than [D-Ser^2, Leu^5]enkephalyl-Thr (DSLET).

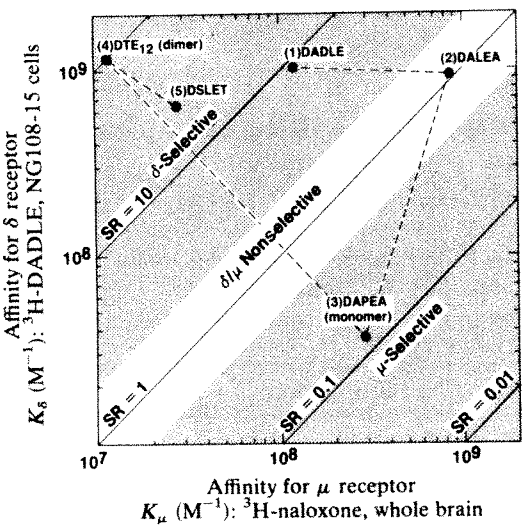


Fig. 1 Two-dimensional display of affinities of enkephalin peptides in a μ-specific radioligand assay (K_μ, abscissa), and in the δ-specific radioligand assay (K_δ, ordinate). Non-selective ligands would fall in the region along the main diagonal, μ-selective ligands below and to the right, and δ-selective ligands above and to the left. Compare (1) DADLE, the δ prototype; (2) DALEA, the amidated pentapeptide enkephalin; (3) DAPEA, the amide of the tetrapeptide [D-Ala^2, des-Leu^5]enkephalin; (4) DTE$_{12}$, the dimer of DAPEA, connected by a methylene bridge with $n = 12$; and (5) DSLET, the best δ-selective ligand previously reported. Fine diagonal lines correspond to 10-fold changes in the δ/μ SR. All affinities were calculated using the formula: $K_1/K_2 = (IC_{50})_2/(IC_{50})_1$, which is valid in the present experimental conditions with low per cent binding of labelled ligand[19]. The calculations used the following data, obtained in similar conditions: in the δ activity assay, DADLE shows $K = 1.03 \times 10^9$ M^{-1} (N = 4 expts) and an $IC_{50} = 1.15 \pm 0.07$ nM ($N = 4$); in the μ assay, naloxone displays $K = 6.6 \pm 1.8 \times 10^8$ M^{-1} ($N = 6$) and $IC_{50} = 1.56 \pm 0.15$ nM ($N = 3$).

In the 'μ receptor' assay, when ^{3}H-NAL was used as labelled ligand, DAPEA monomer retained high affinity, was only three-fold less potent than DALEA, and was twice as potent as DADLE and 10-fold more potent than DSLET. Dimers with $n = 2$–8 (DTE$_2$–DTE$_8$) showed higher potencies than the monomer DAPEA, and had IC_{50} values of 1.8–2.7 nM. DTE$_{10}$ was slightly less active than DAPEA. Surprisingly, the dimer of tetrapeptide enkephalins with $n = 12$ (DTE$_{12}$) showed an extremely low affinity for μ receptors: approximately 28-, 12- and 80-fold lower affinity than DAPEA, DADLE and DALEA respectively.

We may define the δ/μ selectivity ratio (SR) as the ratio of the IC_{50} using ^{3}H-NAL and whole brain divided by the IC_{50} obtained using ^{3}H-DADLE and NG108-15 cells (Table 1). A completely non-selective compound with the same potency in both assays should have a ratio of unity[9]. For example, DALEA shows a SR of 0.97, and thus serves as a non-selective ligand. The tetrapeptide DAPEA shows a SR of 0.1, corresponding to its 10-fold preference for μ receptors, while DADLE (SR = 7.2) and DSLET (SR = 20) are δ-selective peptides. Dimers DTE$_2$ and DTE$_4$ are non-selective (SR ~ 1), while DTE$_6$ and DTE$_8$ slightly favour the μ and δ receptor sites respectively. The dimer DTE$_{10}$, with a longer cross-linking methylene chain, favours δ receptors (SR = 4.7). The dimeric tetrapeptide enkephalin DTE$_{12}$ shows an extraordinary selectivity for δ receptors: its SR of 91 is approximately 13 times that of DADLE and 4.5 times that of DSLET, which has been regarded as the most δ-selective ligand available to date. The DTE$_{12}$ dimer shows a 30-fold greater affinity for the δ receptor and a 30-fold weaker affinity for the μ receptor, relative to the monomer DAPEA.

The inter-relationships among the various ligands are illustrated in Fig. 1. The abscissa indicates the equilibrium constant of association (affinity constant K_μ) estimated using the μ-specific radioligand assay. The ordinate shows the affinity (K_δ) in the δ-specific radioligand assay. The line of identity corresponds to a δ/μ non-selective ligand. The region below the line of identity corresponds to a μ-selective ligand, that above the line to δ-selectivity. Lines corresponding to 10-fold changes in SR are shown. DADLE (point number 1) is the progenitor of the δ-specific ligand, with a K_δ of $\sim 10^9$ but a K_μ of 10^8. Amidation of [D-Ala^2, Leu^5]enkephalin, forming DALEA (point 2), results in loss of selectivity. The amide of the tetrapeptide (DAPEA, point 3) shows ~10-fold selectivity for the μ receptor, consistent with previous reports[14,15]. However, the dimer of this tetrapeptide with a methylene bridge of $n = 12$ (DTE$_{12}$, point 4) shows a dramatic increase in δ selectivity ($K_\delta = 10^9$, $K_\mu = 10^7$). Its affinity and selectivity surpass those of DSLET (point 5).

The present results for DADLE and DSLET are consistent with those derived from *in vitro* muscle bioassays. Gacel *et al.*[18] reported that DSLET was seven times more selective than DADLE when evaluated in terms of its inhibitory potencies on the electrically stimulated mouse vas deferens and guinea pig ileum (corresponding to δ and μ bioassays respectively).

The present findings (Table 1, Fig. 1) are consistent with the hypothesis that the dimers of the tetrapeptide can serve as bivalent ligands, binding simultaneously to two distinct but closely clustered δ receptors, but failing to bridge two μ receptors, and complement those for the dimers of the pentapeptide leucine enkephalins, (H-Tyr-D-Ala-Gly-Phe-Leu-NH-)$_2$ (-CH_2-)$_n$. If the C-terminal leucine residue were approximately as long as a methylene chain with $n = 3$, then chain lengths of $n = 2$–6 for the dimeric pentapeptides would correspond to $n = 8$–12 for the dimeric tetrapeptides. Thus it is not surprising that the most selective and potent tetrapeptide has the longest methylene bridge, whereas the most potent pentapeptide has

Table 1 Receptor binding activities of enkephalin analogues

Enkephalins	IC_{50} (nM) δ Activity (^{3}H-DADLE)	μ Activity (^{3}H-NAL)	δ/μ Selectivity ratio $(IC_{50})_\mu/(IC_{50})_\delta$
Monomer			
DAPEA: H-Tyr-D-Ala-Gly-Phe-NH_2	33.2	3.48	0.10
DALEA: H-Tyr-D-Ala-Gly-Phe-Leu-NH_2	1.24	1.20	0.97
DADLE: H-Tyr-D-Ala-Gly-Phe-D-Leu-OH	1.15	8.22	7.2
DSLET: H-Tyr-D-Ser-Gly-Phe-Leu-Thr-OH	1.85	37.9	20
Dimers; DTE_n (H-Tyr-D-Ala-Gly-Phe-NH–$(CH_2)_n$–NH-Phe-Gly-Ala-D-Tyr-H)			
DTE_2	1.68	1.81	1.1
DTE_4	2.87	2.79	0.97
DTE_6	3.67	2.28	0.62
DTE_8	1.46	2.66	1.8
DTE_{10}	1.04	4.84	4.7
DTE_{12}	1.06	96.3	91

Results are expressed as the IC_{50} for each compound in an assay for δ activity using ^{3}H-DADLE (0.10 nM) and NG108-15 cell membranes according to Chang *et al.*[5,16], and in an assay for μ activity using ^{3}H-naloxone (0.15 nM) and membranes prepared from whole brain[17]. Specific binding was measured as the difference between total binding and that in the presence of 10^{-5} M DADLE or 10^{-5} M naloxone. Competition curves were analysed by computerized non-linear least-squares estimation of the logistic curves, to obtain an estimate of the IC_{50} with its standard error[20]. IC_{50} values showed a within-experiment coefficient of variation (c.v.) of 6%, and a between-experiment c.v. of 14% for the δ activity assay, with values of 14 and 16% respectively in the μ activity assay. The δ/μ selectivity ratio is defined as the IC_{50} in the μ assay divided by the IC_{50} in the δ assay.

the shortest methylene bridge ($n = 2$): the observations for both series of dimers are consistent with very close spacing of δ receptors.

To establish that the special properties of the dimer is due to its bivalency, we are currently synthesizing alkylamidated derivatives of the enkephalin monomers and 'handicapped dimers' with mono-*N*-acetylation of one of the tyrosines in both the tetra- and pentapeptide series. Studies of these new compounds should help to exclude alternative hypotheses to the cross-linking mechanism. Presumably, the lack of increased affinity of DTE_n or DPE_n for the μ receptor is due to a different spatial pattern of μ receptors in the membrane. This rationale has provided a basis for the design and synthesis of what appear to be superior probes of the δ receptor. A similar strategy should be useful in designing ligands with improved affinity and selectivity for a wide variety of other receptor systems for drugs, hormones and neurotransmitters.

We thank Drs H. Fales and R. D. Macfarlane for mass spectrometry of peptides, and Dr S. Matsuura and P. J. Munson for many helpful discussions. Drs K. J. Catt and S. Krumins provided constructive reviews of the manuscript, and G. M. Thornton provided secretarial assistance.

Received 5 January; accepted 25 March 1982.

1. Martin, W. R., Eades, C. G., Thompson, J. A., Huppler, R. E. & Gilbert, P. E. *J. Pharmac. exp. Ther.* **197**, 517–532 (1976).
2. Lord, J. A. H., Waterfield, A. A., Hughes, J. & Kosterlitz, H. W. *Nature* **267**, 495–499 (1977).
3. Simantov, R., Childers, S. R. & Snyder, S. H. *Eur. J. Pharmac.* **47**, 319–331 (1978).
4. Robson, L. E. & Kosterlitz, H. W. *Proc. R. Soc.* **B205**, 425–432 (1979).
5. Chang, K.-J. & Cuatrecasas, P. *J. biol. Chem.* **254**, 2610–2618 (1979).
6. Wuster, M., Schulz, R. & Herz, A. *Life Sci.* **27**, 163–170 (1980).
7. Hazum, E., Chang, K.-J. & Cuatrecasas, P. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3038–3041 (1980).
8. Goodman, R. R., Snyder, S. H., Kuhar, M. J. & Young, W. S. III *Proc. natn. Acad. Sci. U.S.A.* **77**, 6239–6243 (1980).
9. Kosterlitz, H. W., Lord, J. A. H., Paterson, S. J. & Waterfield, A. A. *Br. J. Pharmac.* **68**, 333–342 (1980).
10. Rónai, A. Z. *et al. Eur. J. Pharmac.* **69**, 263–271 (1981).
11. Costa, T., Shimohigashi, Y., Matsuura, S., Chen, H.-C. & Rodbard, D. in *Peptides; Proc. 7th Am. Peptide Symp.* (eds Rich, D. H. & Gross, E.) 625–628 (Pierce Chemical, Rockford, 1981).
12. Shimohigashi, Y., Costa, T., Matsuura, S., Chen, H.-C. & Rodbard, D. *Molec. Pharmac.* **21**, 558–563 (1982).
13. Snyder, S. H. *Science* **209**, 976–983 (1980).
14. Rónai, A. Z., Székely, J. I., Berzétei, I., Miglécz, E. & Bajusz, S. *Biochem. biophys. Res. Commun.* **91**, 1239–1249 (1979).
15. McGregor, W. H., Stein, L. & Belluzzi, J. D. *Life Sci.* **23**, 1371–1378 (1978).
16. Chang, K.-J., Miller, R. J. & Cuatrecasas, P. *Molec. Pharmac.* **14**, 961–970 (1978).
17. Pert, C. B. & Snyder, S. H. *Molec. Pharmac.* **10**, 868–879 (1974).
18. Gacel, G., Fournie-Zaluski, M. C., Fellion, E. & Roques, B. P. *J. med. Chem.* **24**, 1119–1124 (1981).
19. Munson, P. J. & Rodbard, D. *Analyt. Biochem.* **107**, 220–239 (1980).
20. DeLean, A., Munson, P. J. & Rodbard, D. *Am. J. Physiol.* **235**, E97–E102 (1978).

Characterization of the structural gene and putative 5′-regulatory sequences for human proopiomelanocortin

Madeleine Cochet*, Annie C. Y. Chang & Stanley N. Cohen

Departments of Genetics and Medicine, Stanford University School of Medicine, Stanford, California 94305, USA

It is now well established that the hormones corticotropin (ACTH) and β-lipotropin (β-LPH) are contained within a common precursor peptide called proopiomelanocortin (POMC), which is synthesized in the anterior and/or intermediate lobes of the pituitary[1–3] and in the hypothalamus[4]. The recent cloning of a 'full-size' cDNA copy of bovine POMC mRNA has revealed the presence of additional peptides in the previously 'cryptic' segment of the precursor protein[5]. This cDNA has been used as a hybridization probe for the isolation and characterization of a bovine genomic DNA segment encoding the complete POMC mRNA[6], and for isolation of part of the corresponding human genomic DNA sequence[7]. DNA sequence analysis of these isolates and of part of the rat POMC gene[8] has shown that, in the species studied, certain segments of the POMC structural gene have been conserved during evolution, whereas the nucleotide sequences of other segments have diverged sharply. We have now isolated and characterized a 11.6-kilobase (kb) human genomic DNA fragment that encodes the complete POMC mRNA and putative regulatory sequences proximate to the gene (which is repressed by glucocorticoids *in vivo*[9]). We have also analysed the 800-base pair (bp) segment preceding the mRNA coding sequence, within which—at a position nearly 480 bp upstream from the mRNA start site—is a 21-bp segment that shares homology with a DNA sequence beginning 370 bp upstream from two other glucocorticoid-controlled genes. We speculate that these sequences may be involved in the regulation of POMC gene expression by glucocorticoids.

* Present address: Unité de Biologie Moléculaire du Gène, Institut Pasteur, 28, Rue du Docteur Roux, Paris 75724, France.

0028-0836/82/210335-05$01.00

Using probes made by nick-translation of segments of a previously characterized human POMC genomic DNA fragment (Fig. 1, probes *a* and *b*)[7,10], we screened a human fetal DNA library (a gift of T. Maniatis) by *in situ* hybridization in standard conditions (Fig. 1 legend). Two bacteriophage λ plaques that reacted with probe *a* also hybridized with probe *b*, which only contains sequences corresponding to the 5′ end

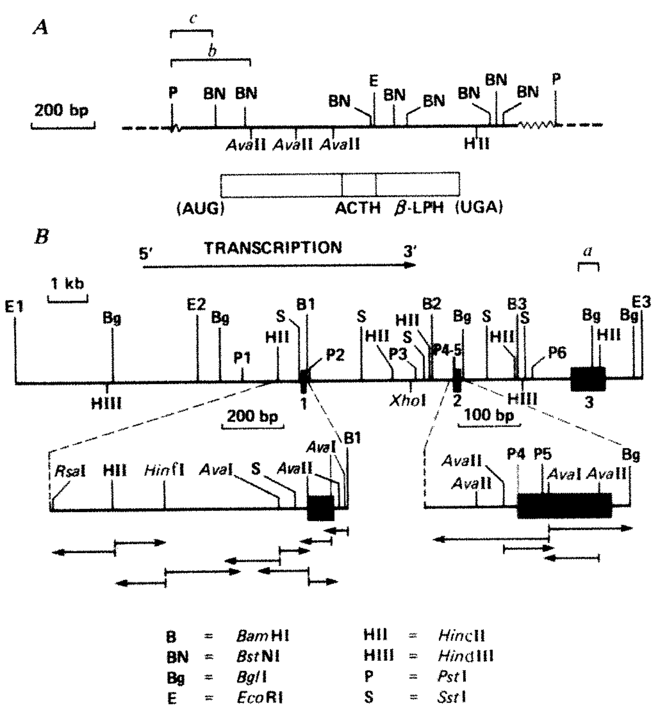


Fig. 1 *A*, partial restriction endonuclease map of the bovine POMC cDNA insert of pSNAC20 (ref. 5). The broken lines represent pBR322 DNA sequences. The serrated lines represent the 27- and 112-nucleotide poly(dG)·poly(dC) tracts lying at junctions between the plasmid vector and the cloned cDNA. The 5′ and 3′ termini of double-stranded cDNA are designated according to the corresponding ends of bovine POMC mRNA. *b* and *c* indicate the probes used for the hybridization studies described in the text and were prepared as follows. An *Ava*II endonuclease-generated fragment of the pSNAC20 plasmid, containing 248 bp of the 5′ end from the bovine cDNA insert plus 119 bp of pBR322 sequences, was inserted into the *Hinc*II cleavage site of pACYC177 by blunt end ligation after filling-in of the ends by DNA polymerase. The resulting plasmid (pACYC654) was digested by the *Pst*I restriction endonuclease, generating two fragments, one of which contains the 5′ end of the bovine cDNA. This fragment (530 bp) was then purified by sucrose gradient centrifugation, and used as probe *b*. Probe *b* was also digested by *Bst*NI endonuclease, giving three fragments, and the resulting 134-bp fragment (which contained most of the sequences corresponding to the 5′-untranslated region of bovine POMC mRNA) was purified by electrophoresis, yielding the *c* probe. The open bar below the POMC cDNA corresponds to the POMC peptide; the initiation and termination codons are in brackets. *B*, partial mapping of the structural gene encoding human POMC mRNA and strategy of sequencing. The human fetal DNA library used was constructed by inserting DNA fragments from a partial *Eco*RI endonuclease digest into a bacteriophage λ charon 4A vector[30]. It was screened by *in situ* hybridization[31] using the nick-translated[7,10] *a* probe. This probe corresponds to an *Ava*I–*Hinc*II fragment of a human genomic DNA segment previously cloned in the pACYC401 plasmid[7] and contains sequences encoding the ACTH–βLPH region of POMC. The partial restriction enzyme cleavage map corresponding to the genomic DNA contained in H-POMC λ5A (one of the two identical clones we obtained) was deduced by analysis of DNA samples singly or doubly digested with restriction endonucleases. The black thick bars and the numbers below them designate DNA sequences corresponding to POMC mRNA (exons). Their positions were determined by hybridization, with the nick-translated probes *a*, *b* or *c*, of endonuclease-treated DNA digests transferred onto nitrocellulose filters[32] or aminothiophenol paper (see Fig. 3 legend). Hybridization conditions with probes *a* and *b* have been described previously[7]; conditions used with probe *c* are described in Fig. 3 legend. For convenience, DNA fragments were identified by the abbreviated names of the endonuclease cleavage sites that bracket them, followed by a number designation according to the direction of transcription (for example, E2–E3 is the DNA fragment terminated at the 5′ end by *Eco*RI cleavage site number 2 and at the 3′ end by *Eco*RI site number 3). The DNA fragments carrying exons 1 and 2 are enlarged in the diagram and the strategy of sequencing is shown. The arrows designate the segments sequenced; the tails indicate the sites of ^{32}P end-labelling of the fragments.

of bovine POMC mRNA. Analysis of the DNA from these clones by restriction endonuclease digestion and electrophoresis on agarose gels indicated that both included *Eco*RI fragments 11.6 and 4.7 kb long.

DNA from one of these clones (H-POMC λ5A) was analysed further by electrophoresis of products resulting from single or double digestion with restriction endonucleases (Fig. 1*B*). Only the E2–E3 segment of the 11.6-kb fragment hybridized to both probes; DNA fragments B3–E3 and P6–E3 hybridized only to probe *a*, which includes all of the protein-coding sequences for ACTH and β-LPH, whereas DNA fragments B2–B3 and P5–P6 hybridized only to probe *b*. These findings suggested that at least a proportion of the sequences encoding the amino-terminal end of the precursor peptide[5] is contained in fragment B2–B3, and permitted the endonuclease cleavage sites of the genomic DNA fragment to be oriented with respect to transcription (see Fig. 1).

Finer mapping of the B2–B3 fragment following subcloning using the vector pBR322 indicated that sequences homologous to probe *b* are localized to the region surrounding the P5 site (data not shown). The DNA sequence for this region (Fig. 2) was determined according to the strategy shown in Fig. 1*B*. A high degree of homology was observed for 152 bp (nucleotides +88 to +239) in the region designated as exon 2 in Fig. 2, which is contiguous with a previously identified region of interspecies conservation that begins near the 5′ end of exon 3. The conserved region is delineated by intron–exon junction consensus sequences[11], which include G-T at the 5′ end and A-G at the 3′ end of the intron. Exon 2 contains the AUG translational start codon for the POMC peptide, the conserved downstream codons specifying the hydrophobic leader sequences of human and bovine POMC, and only a short (20 bp) segment of the untranslated region of POMC mRNA. Thus, the remainder of the untranslated region was either not present on the 11.6-kb fragment, or alternatively, the lack of detectable hybridization of probe *b* with any other segment of this fragment might be explained by evolutionary divergence in the sequences of the two mammalian genes.

As human POMC mRNA was not readily available, we re-probed the 11.6-kb fragment with the 5′ end of bovine DNA, using less stringent hybridization conditions that might allow identification of regions of incomplete homology. Fragments of H-POMC λ5A DNA generated by *Eco*RI and *Bam*HI endonucleases were subcloned, transferred to aminothiophenol paper (B. Seed, personal communication) after electrophoresis in gels, and hybridized at various temperatures between 45 and 65 °C with probe *c*, which contains 107 bp of the extreme 5′ end of bovine POMC cDNA (15 bp of which are included in exon 2). At 56 °C, the temperature that gave the best signal over background, a positive signal was obtained only with the genomic DNA fragment containing exon 2 (Fig. 3, lane 3) and with the E2–B1 fragment (lane 2). More detailed analysis of fragment E2–B1 (Fig. 3*b*) localized the homologous sequence to a 160-bp *Ava*I endonuclease-generated fragment near the *Bam*HI terminus of the E2–B1 fragment (Fig. 3*b*, lanes 3, 4). The DNA sequence of this fragment and the surrounding regions between the *Rsa*I and *Bam*HI sites was determined according to the strategy shown in Fig. 1*B*.

Alignment of the nucleotide sequence (Fig. 2) with the published sequence for bovine POMC genomic DNA[6] (starting at position −181) allows assignment of a putative start site for the human mRNA (+1). The dinucleotide G-T is found 87 bp downstream from this start site at the end of a region that contains three segments showing >80% homology with the corresponding bovine sequences (positions +9 to +28, +43 to +63 and +71 to +87). We conclude therefore that this region may contain the first mRNA-coding segment (that is, exon 1) of the human gene, despite its different length from the corresponding bovine segment (87 instead of 108 bp). A 'TATA' box characteristic of an RNA polymerase II (or B) promoter site[12] and a 'CAAT' box[13] are found 28 and 63 bp respectively upstream from the putative start site of the POMC mRNA at

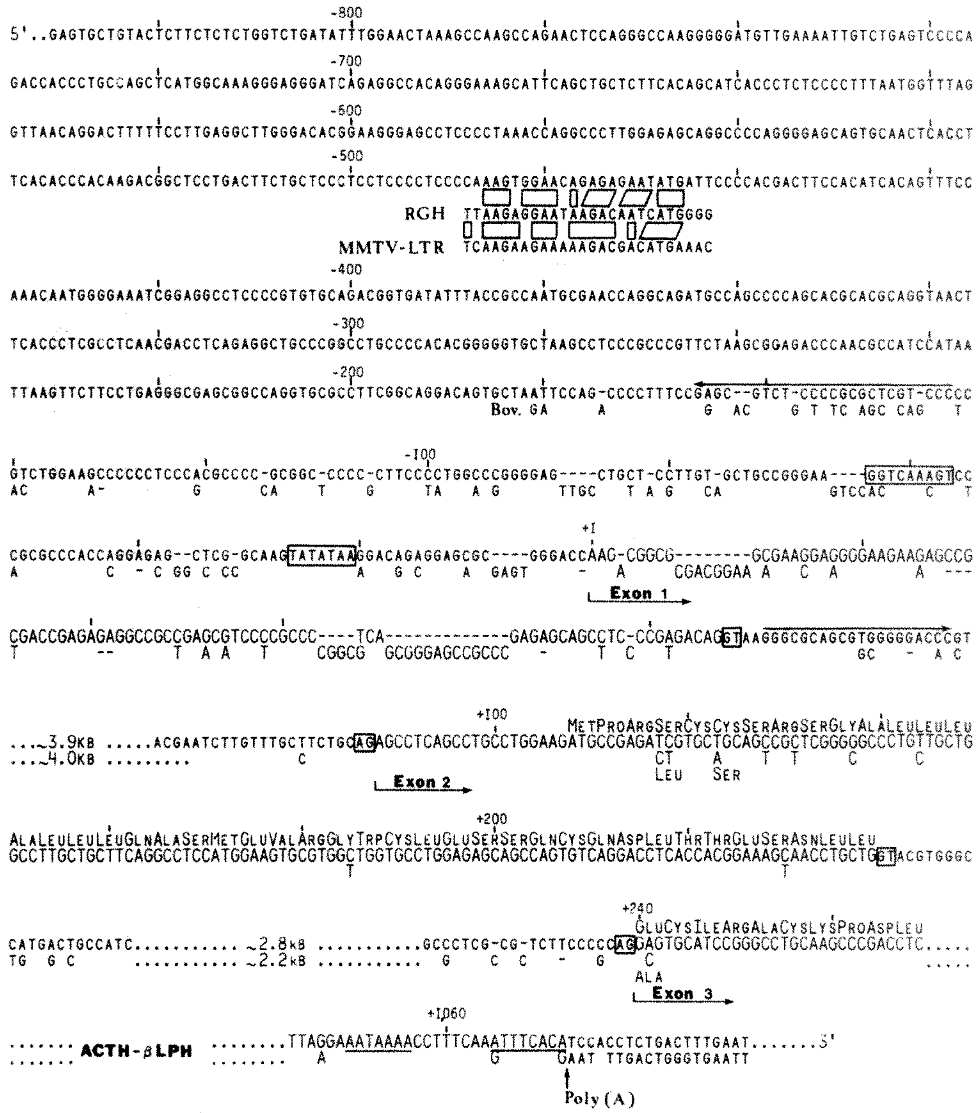


Fig. 2 Nucleotide sequence of the human POMC genomic DNA strand corresponding to POMC mRNA. Digests of DNA fragments E2–B1 and B2–B3, which had been subcloned using the pBR322 plasmid vector, were digested with *Ava*I, *Ava*II, *Bam*HI or *Hin*fI endonucleases and treated with polynucleotide kinase[33]. The nucleotide sequence was then determined[33] by following the strategy of Fig. 1*B*. The sequence of the strand that corresponds to mRNA is shown. Differences in nucleotides and amino acids in the bovine versus human sequences are indicated below the human DNA sequence. Broken lines represent nucleotides absent from the genomic DNA sequence. Position +1 indicates the putative start site of human POMC mRNA based on known bovine sequences[6]. Nucleotides are numbered plus for exonic sequences downstream from the mRNA start site and minus for sequences upstream from the mRNA start site. The TATA and CAAT sequences, and the A-G, G-T dinucleotides at the intron–exon junctions are boxed. The proposed consensus sequence[13] for the poly(A) addition site is underlined. The region of dyad symmetry is overlined by arrows. The region of homology between human DNA sequences and RGH[22] or MMTV-LTR[23] are shown by boxes; in both RGH and MMTV-LTR, the homologous sequence begins 392 nucleotides upstream from the mRNA start site. Errors in the partial sequence of exon 3 published previously[7] have been corrected as a result of additional analysis, and the sequence determined agrees with the DNA sequence of Takahashi *et al.*[27] and the protein sequence of Seidah and Chrétien[28,29]. Despite careful sequencing and reading we were unable to decide whether nucleotides −164 and 165 are G-A or A-G.

positions similar to those described for other mammalian genes[14]. Upstream from the CAAT box of the human POMC gene are runs of nucleotides (positions −64 to −75, −101 to −145, −165 to −179) that are significantly homologous with sequences found in the corresponding positions of the bovine gene; at other locations the bovine and human sequences are highly divergent. Interestingly, a region of very stable dyad symmetry ($\Delta G = -47$ kcal) is located 100 bp upstream from the CAAT box and at the 3′ end of intron A, therefore enclosing the CAAT and TATA boxes and all of exon 1.

If the identifications made above on the basis of DNA sequence analysis are correct, the E2–B1 segment of the 11.6-kb DNA fragment cloned in H-POMC λ5A should contain a promoter for RNA polymerase II. We demonstrated the presence of such a promoter in this region by *in vitro* transcription experiments[15–17] using endonuclease-generated digests of cloned POMC human genomic DNA (plasmids pλP2 and pλEB7) as templates for RNA polymerase II. The plasmids used for these studies contained respectively the human genomic DNA fragments P1–P2 and E2–B1; if the regions identified as TATA and CAAT boxes in these clones are in fact signals allowing initiation of trancription of the POMC gene, digests of pλP2 DNA by *Bam*HI, *Pst*I, or *Pvu*I endonuclease should yield *in vitro* RNA 'run off'[15] transcripts of 130, 300, and 427 nucleotides respectively, and a digest of pλEB7 DNA by the *Sal*I endonuclease should give a 405-nucleotide RNA transcript. The RNA transcripts synthesized *in vitro* on cloned POMC genomic DNA have the lengths predicted (Fig. 4). Thus, the sequences we have identified as the putative transcriptional regulatory region for the POMC gene seem capable of directing RNA polymerase II-dependent initiation of transcription *in vitro*.

Much work has focused on the role in gene regulation of sequences immediately upstream from the mRNA-coding regions of prokaryotic[18] and eukaryotic[14] genes, but little attention has been given to the possible influence of more distant upstream sequences. Recently, however, a putative binding sequence for the progesterone–receptor complex involved in the control of egg-white protein synthesis in the chicken oviduct has been identified 250–300 bp upstream from the mRNA start site of the ovalbumin gene[19]. In the case of the POMC gene, sequences at an equal or greater distance from the mRNA start site might be important in the glucocorticoid regulation of gene expression, or in the abnormal production of the POMC-derived peptides by certain non-pituitary tumours[20,21]. To explore the role of more-distant upstream regions in normal and abnormal POMC gene expression, we have determined the primary sequence of the 800-nucleotide segment preceding the POMC mRNA start site.

Comparison of the DNA sequence obtained by such analysis with published sequences of other glucocorticoid-controlled genes showed that between position −486 and −466 upstream

from the POMC mRNA start site is a 21-bp DNA segment (see Fig. 2) that shows 80% homology with a 21-bp stretch beginning 370 bp upstream from the mRNA start site of the rat growth hormone (RGH) gene[22]. A similar sequence (75% homology) starts at the same location relative to the mRNA start site of the long terminal repeat (LTR) of the mouse mammary tumour virus (MMTV) DNA[23]. The expression of both the RGH gene and the MMTV genetic sequence are positively regulated by glucocorticoids[24,25], whereas the expression of the POMC gene is negatively regulated by this hormone[9].

The existence of specific DNA sequences that may be involved in the binding of the hormone-receptor complexes has been suggested recently by *in vitro* studies for the MMTV LTR gene[26] and the egg-white protein genes in the chicken oviduct[19]. The presence of largely homologous sequences in two gene segments positively controlled by glucocorticoids, and also in

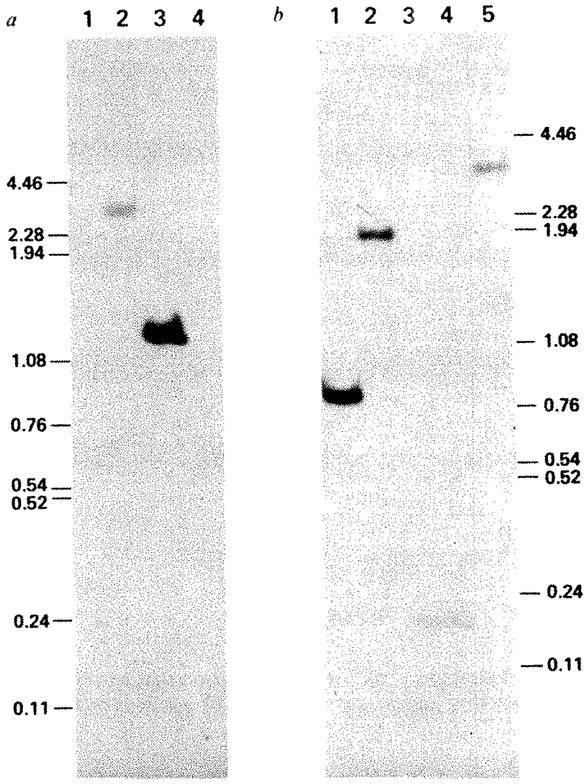


Fig. 3 Detection of the 5'-untranslated region of human POMC mRNA. Genomic DNA fragments E1–E2, E2–B1, B1–B2 and B2–B3 generated by *Eco*RI and *Bam*HI restriction endonuclease digestion of H-POMC λ5A DNA were subcloned using the pBR322 vector. The DNA of plasmids containing the various fragments was digested by restriction endonucleases, subjected to electrophoresis on 2% agarose gels, transferred on aminothiophenol paper (B. Seed, personal communication) and hybridized to 20 ng of the nick-translated probe *c* (10^8 c.p.m. μg^{-1}) for 48 h at 56 °C in 5×SSC, 0.1 M $NaPO_4$ *p*H 6.8, 10× Denhardt's solution[34]. After hybridization, the filters were washed at room temperature for 10 min in 2×SSC, 0.1% SDS twice, then 10 min in 0.5×SSC, 0.1% SDS once, and subjected to autoradiography for 12 h at −70 °C with XR-5 film and Dupont Cronex lightening plus intensifying screen. *a*: Lane 1, 0.4 μg of E1–E2 fragment digested with *Hin*dIII endonuclease; lane 2, 0.4 μg of E2–B1 fragment; lane 3, 0.5 μg of B2–B3 fragment digested by *Sst*I; lane 4, 0.5 μg of B1–B2 fragment. *b*, Digests of 0.5 μg E2–B1 fragment with *Hin*cII endonuclease (lane 1), *Pst*I (lane 2), *Ava*I (lane 3), *Ava*I plus *Pst*I (lane 4). Lane 5 contains undigested DNA. The numbers at the side of the figure indicate the molecular size (in kb) of marker DNA derived from λ wild-type DNA digested by *Hin*dIII, and SV40 DNA digested by *Hin*fI endonuclease. The hybridization signal seen with probe *c* (*a*, lane 3) is more intense than expected. We attribute this effect to slight contamination of probe *c* with probe *b* sequences, despite electrophoretic purification of the *c* probe fragment before use.

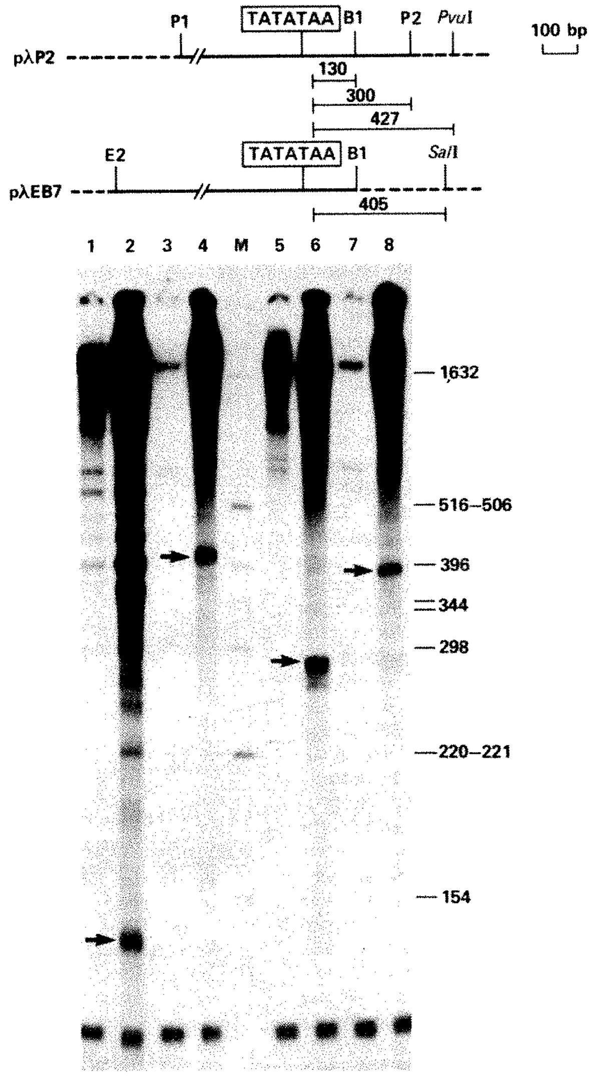


Fig. 4 Run-off analysis of RNA synthesized *in vitro* on human genomic DNA by RNA polymerase II (or B). The DNA fragments E2–B1 and P1–P2 of clone H-POMC λ5A were inserted into the pBR322 vector, yielding the chimaeric plasmids pλEB7 and pλP2. The resulting DNAs were used as templates for *in vitro* transcription by cell-free extracts of HeLa cells (BRL) as described[17]. The DNA fragments were either excised and purified by sucrose gradient centrifugation (fragment P1–P2, lanes 5, 6) or treated with restriction endonuclease and phenol-extracted (lanes 1–4, 7, 8). To differentiate transcripts made by RNA polymerase II (or B) from those resulting from RNA polymerase III (or C), 1 $\mu g\ ml^{-1}$ of α-amanitine was added to appropriate reaction mixtures (lanes 1, 3, 5, 7). RNA transcripts of the human genomic DNA were purified as described previously[16] and analysed on a 4.5% acrylamide/7 M urea gel. After drying, the gel was exposed to film for 6 h as described in Fig. 3 legend. The length of the RNA transcripts was determined relative to molecular length markers (lane M) consisting of fragments of pBR322 DNA produced by digestion with *Hin*fI and ^{32}P end-labelled. The lengths (in numbers of nucleotides, nt) are labelled on the right side of the panel. The region of initiation of transcription was inferred from the length of the RNA chains which are terminated prematurely by endonuclease digestion of the template DNA at specific sites, and which therefore 'run-off' the template[15]. The arrows point to the RNA transcripts described. Lanes 1–4, run-off transcripts of 4.5 μg of $p\lambda P_2$ double-digested with *Pst*I and *Bam*HI (lanes 1, 2) or *Pvu*I alone (lanes 3, 4). The α-amanitine-sensitive transcript has an apparent length of 135 (lane 2) and 415 (lane 4) nt. Lanes 5, 6, run-off transcripts of 2.8 μg of the excised P1–P2 fragment. The α-amanitine transcript has an apparent length of 292 nt (lane 6). The molecular weight was calculated for the upper transcript. The smear below is probably due to degradation because it does not appear each time. Lanes 7, 8, run-off transcript of 4 μg of pλEB7 digested by *Sal*I. The DNA transcript has an apparent length of 390 nt (lane 8).

the region upstream from the negatively controlled human POMC gene, could be accounted for if such sequences are involved in the binding to DNA of the glucocorticoid–receptor complexes. Although the homology in the DNA segments we have identified is incomplete, the occurrence of similar sequences at the same distance upstream from the mRNA start site in RGH and the MMTV LTR is of interest, although their functional significance remains unknown. In any case, the specific DNA sequence we have identified does not seem to be necessary for dexamethasone stimulation of MMTV LTR-regulated gene expression in transfected cells: stimulation still occurs after removal of all MMTV LTR sequences beyond the 170–200 nucleotides immediately upstream from the mRNA cap site (J. Majors and H. Varmus, and G. Ringold, personal communications).

Our data indicate that the 11.6-kb *Eco*RI endonuclease-generated gene fragment E2–E3 includes the entire structural gene for human POMC. The gene's mRNA coding sequences are contained in three exons of 87, 152 and 833 bp, separated by two introns of about 3.9 and 2.8 kb. While segments of both interspecies conservation and divergence exist in the mRNA-coding and non-coding regions of the human, bovine and rat POMC genes, the mRNA of all three species is spliced at identical sites. The longest segments of interspecies conservation occur in the protein-coding region of the gene, especially in the segment that includes the hormones ACTH and β-LPH. However, significant sequence conservation occurs also in the 3′-non-coding region and in a segment upstream from the TATA and CAAT boxes (−64 to −178), suggesting that these regions may be important in the expression of POMC *in vivo*.

After completing this work, we received a manuscript reporting a similar overall structure and DNA sequence of the mRNA-encoding region of the POMC gene and the immediately proximate upstream region[27]. The protein sequence recently reported in refs 28 and 29 corresponds to the DNA sequence that both we and Takahashi *et al.* have obtained.

These studies were supported by the US NIH. M.C. was a postdoctoral research fellow of the CNRS.

Received 3 August 1981; accepted 31 March 1982.

1. Mains, R. E., Eipper, B. A. & Ling, N. *Proc. natn. Acad. Sci. U.S.A.* **74,** 3014–3018 (1977).
2. Roberts, J. L. & Herbert, E. *Proc. natn. Acad. Sci. U.S.A.* **74,** 4826–4830 (1977).
3. Nakanishi, S., Inoue, A., Taii, S. & Numa, S. *FEBS Lett.* **84,** 105–109 (1977).
4. Krieger, D. T. & Liotta, A. S. *Science* **205,** 366–372 (1979).
5. Nakanishi, S. *et al. Nature* **278,** 423–427 (1979).
6. Nakanishi, S. *et al. Eur. J. Biochem.* **115,** 429–438 (1981).
7. Chang, A. C. Y., Cochet, M. & Cohen, S. N. *Proc. natn. Acad. Sci. U.S.A.* **77,** 4890–4894 (1980).
8. Drouin, J. & Goodman, H. M. *Nature* **288,** 610–612 (1980).
9. Nakanishi, S., Kita, T., Taii, S., Imura, H. & Numa, S. *Proc. natn. Acad. Sci. U.S.A.* **74,** 3283–3286 (1977).
10. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. *J. molec. Biol.* **113,** 237–251 (1977).
11. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. & Chambon, P. *Proc. natn. Acad. Sci. U.S.A.* **75,** 4853–4857 (1978).
12. Goldberg, M. thesis, Stanford Univ. (1979).
13. Benoist, C., O'Hare, K., Breathnach, R. & Chambon, P. *Nucleic Acids Res.* **8,** 127–142 (1980).
14. Breathnach, R. & Chambon, P. *A. Rev. Biochem.* **50,** 349–383 (1981).
15. Weil, P. A., Luse, D. S., Segall, J. & Roeder, R. G. *Cell* **18,** 469–484 (1979).
16. Wasylyk, B., Kédinger, C., Corden, J., Brison, O. & Chambon, P. *Nature* **285,** 367–373 (1980).
17. Manley, J. F., Fire, A., Cano, A., Sharp, P. A. & Gefter, M. L. *Proc. natn. Acad. Sci. U.S.A.* **77,** 3855–3859 (1980).
18. Rosenberg, M. & Court, D. *A. Rev. Genet.* **13,** 313–354 (1979).
19. Mulvihill, E. R., Lepennec, J.-P. & Chambon, P. *Cell* **28,** 621–632 (1982).
20. Wolfsen, A. R. & Odell, W. D. *Am. J. Med.* **66,** 765–772 (1979).
21. Imura, H. *Adv. Cancer Res.* **33,** 39–75 (1980).
22. Barta, A., Richards, R. I., Baxter, J. D. & Shine, J. *Proc. natn. Acad. Sci. U.S.A.* **78,** 4867–4871 (1981).
23. Donehower, L. A., Huang, A. L. & Hager, G. L. *J. Virol.* **37,** 226–238 (1981).
24. Baxter, J. D. *et al. Cold Spring Harb. Conf. Cell Proliferation* Vol. 6, 317–337 (Cold Spring Harbor Laboratory, New York, 1979).
25. Ringold, G. M., Yamamoto, K. R., Bishop, J. M. & Varmus, H. E. *Proc. natn. Acad. Sci. U.S.A.* **74,** 2879–2883 (1977).
26. Payvar, F. *et al. Proc. natn. Acad. Sci. U.S.A.* **78,** 6628–6632 (1981).
27. Takahashi, H., Teranishi, Y., Nakanishi, S. & Numa, S. *FEBS Lett.* **135,** 97–102 (1981).
28. Seidah, N. G. & Chrétien, M. *Proc. natn. Acad. Sci. U.S.A.* **78,** 4236–4240 (1981).
29. Hsi, K. L., Seidah, N. G., Lu, C. L. & Chrétien, M *Biochem. biophys. Res. Commun.* **103,** 1329–1335 (1981).
30. Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G. & Maniatis, T. *Cell* **15,** 1157–1174 (1978).
31. Benton, W. D. & Davis, R. W. *Science* **196,** 180–182 (1977).
32. Southern, E. M. *J. molec. Biol.* **98,** 503–517 (1975).
33. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65,** 499–560 (1980).
34. Denhardt, D. *Biochem. biophys. Res. Commun.* **23,** 641–646 (1966).

G inversion in bacteriophage Mu: a novel way of gene splicing

Micheline Giphart-Gassler, Ronald H. A. Plasterk & Pieter van de Putte

Laboratory of Molecular Genetics, Leiden State University, Wassenaarseweg 64, 2333 AL Leiden, The Netherlands

The G region of bacteriophage Mu is a 3,000-base pair (bp) invertible DNA segment that determines the host specificity of the phage[1,2]. Only phage with the G segment in the orientation designated (+) can adsorb to *Escherichia coli* K-12 (refs 3–5), whereas phage with the G segment in the (−) orientation infects other hosts such as *Citrobacter freundii* and *E. coli* C (ref. 1). It has been proposed that four genes are involved in host specificity—*S, U, S′* and *U′*—of which *S* and *U* are expressed in the orientation (+) and *S′* and *U′* in the inverted orientation (−) of the G segment. We have presented two alternatives for the organization of these genes (Fig. 1). In model A the four genes involved are entirely contained within the G segment, whereas in the alternative model (B) it is assumed that the proximal part of the *S* gene is located in the α part of the Mu genome. We report here that the *S* and *S′* genes are partially located outside the G segment and share a common part located in α (Sc). This common S region will be joined by G inversion to one of the two variable parts (Sv or Sv′) which results in the synthesis of either tail fibre protein S or S′.

In both models A and B two predictions were made concerning the *S′* gene: (1) that the structural polypeptide with a molecular weight (Mr) of 48,000 (48K) (see Fig. 2*a*,) which is a component of G(−) phage is indeed the product of a gene located at least partially in G and (2) that this *S′* gene is located in the region of G which is non-essential for G(+) phage[6]. Three Mu amber mutants were isolated which were amber phage on G(−) hosts but wild type on the G(+) host *E. coli* K-12. These amber mutations mapped in the G segment in that part which

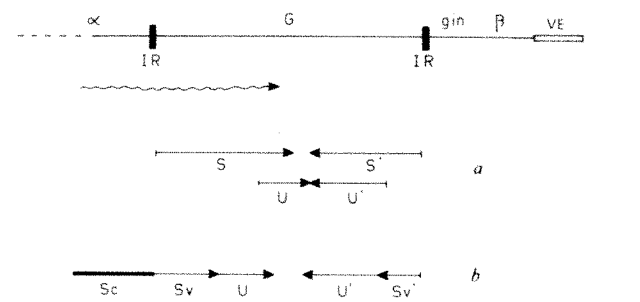


Fig. 1 Two proposed models for the organization of the G region of bacteriophage Mu. The G segment (3 kb) is situated between the α part (~30 kb) and the β part (1.5 kb). The G segment (drawn in orientation +) can invert by recombination between inverted repeats (IR). Amber mutations in genes *S* and *U* have been mapped in G with *U* mutations mapping distal to the *S* mutations[12] and these genes should be located in the leftmost 1,700 bp of G DNA[6]. *S* and *U* are essential for G(+) but not for G(−) phage[1]. The S protein (56K) and a 21K protein (U?) are structural components of G(+) phage particles but not of G(−) phage particles[1,13]. G(−) phage particles contain two structural proteins of 48K and 26K, tentatively designated S′ and U′, respectively, which do not occur in G(+) phage particles. It is assumed that S′ and U′ are involved in adsorption of G(−) phages and that their genes are located in the rightmost 1,300 bp of G DNA non-essential for G(+) phages. To obtain expression of S′ and U′ in the G(−) orientation, transcription of G genes starts in the α part of the Mu genome. The four proteins S, U, S′ and U′ require a coding capacity over 4 kb of DNA, a length exceeding the 3 kb G segment. If the four genes are located entirely in the G segment they should overlap (model A). In model B the *S* and *S′* genes span the α–G boundary with a common region in the α segment (Sc) and two variable regions at the ends of the G segment (Sv and Sv′).

0028–0836/82/210339—04$01.00

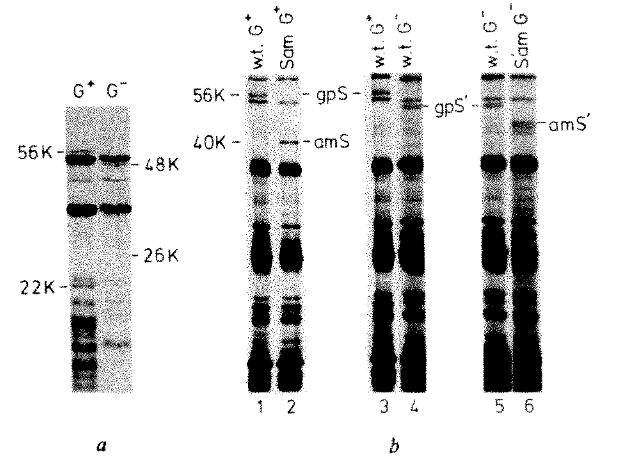


Fig. 2 *a*, Electrophoretic separation of the polypeptides of Mu G(+) and Mu G(−) phage particles[1]. The various structural polypeptides indicated are presumed to represent the following gene products (gp): 56K = gp *S*; 22K = gp *U*; 48K = gp *S'*; 26K = gp *U'*. *b*, Identification of *S* and *S'* gene products in *E. coli* minicells. Minicells were infected with Mu and Mu amber mutants. One hour after infection the minicells were labelled with ^{35}S-L-methionine and the proteins analysed on a 11.5–20% polyacrylamide gradient gel[13]. A Mu polypeptide of M_r 56,000 (56K, lane 1) is not synthesized in the minicells infected with Mu *Sam* 1004 (lane 2). The minicell producing strain PC2251, which resulted from a cross between an *E. coli* C donor and the minicell-producing Adler strain (P678–54), was found to be sensitive for G(−) phages. Lane 4 shows the polypeptide pattern of minicells of this strain infected with G(−) phages, compared with a pattern obtained from minicells of DS410 infected with Mu G(+) (lane 3). A 50K polypeptide is synthesized only after G(−) infection. This polypeptide is absent in minicells infected with Mu G(−) amber mutant 2101 (lane 6) or infected with G(−) amber mutant 2102 and 2103 (results not shown).

is non-essential for G(+) phage, as marker rescue of these mutations was observed with recombinant plasmids containing the right part of the G region (pGP202 and 203, Fig. 3*b*). Therefore, we conclude that there is indeed a region in the G segment which is only essential for G(−) phage. The gene product encoded by this region could be identified in minicells. As can be seen in Fig. 2*b*, minicells infected with Mu G(+) phage synthesize a 56K polypeptide identified as the *S* gene product (lane 2). This polypeptide is absent when G(−)-sensitive minicells are infected with G(−) phage, but another polypeptide of M_r 50,000 (50K, lane 4), comparable with that of the structural *S'* polypeptide, is synthesized instead. This polypeptide is not synthesized in minicells infected with any of the three Mu G(−) amber mutants (lanes 5, 6). These results show that the *S'* gene located in G is essential for G(−) phage only and encodes the structural protein unique for G(−) phage.

To analyse the genetic organization of the G region further, we cloned a Mu DNA fragment containing this region in pBR322. In isolates of this plasmid, pGP56 (Fig. 3*a*), the G region is found for 50% in the (+) and for 50% in the (−) orientation. Cells containing pGP56 express the genes *R*, *S* and *S'* as measured by complementation of Mu amber mutants.

To distinguish between models A and B (Fig. 1), we have isolated various Tn*5* insertions in pGP56, constructed various subclones of this plasmid and tested them for complementation and marker rescue of 'G' amber mutants. Fourteen independent Tn*5* insertions were mapped by restriction enzyme analysis in the G region and in regions flanking G (Fig. 3*a*). Figure 3*a* also shows the level of complementation of Mu *S* and Mu *S'* amber mutants in strains (either *E. coli* K-12 or *E. coli* C) harbouring pGP56 with the various Tn*5* insertions. The results allow the following conclusions. (1) Both *S* and *S'* are expressed from a promoter which is located in α between 900 bp (position 22, S and S'(+)) and 350 bp (position 44, S and S'(−)) to the left of the G segment. (2) The entire *S'* gene or Sv' is, as postulated, located in the right part of the G segment (in the (+) orientation; see positions 69 and 43). (3) Two Tn*5* insertions (44 and 48), located outside the G segment in α, give loss of expression of both *S* and *S'*. Therefore, these insertions are either in the common part of *S* and *S'* or are located between the promoter and the translation start of *S* and *S'*. (4) The promoter distal end of the *S* gene seems to be located in the vicinity of the *KpnI* site (975 bp removed from the left inverted repeat), as Tn*5* insertions 66 and 42 allow complementation of Mu *S* amber mutants.

To localize both the proximal and distal ends of *S* and *S'*, marker rescue and complementation experiments were performed with subclones of pGP56 (see Fig. 3*b*). pGP202, which is deleted for the Mu DNA between the two *HpaI* sites in the G region of pGP56, did not complement Mu *S* or *S'* amber mutants. This result indicates that the C-terminal end of both genes *S* and *S'* stretches beyond the two *HpaI* sites, located respectively 600 and 900 bp away from the left and right inverted repeats. Marker rescue of the *Sam*1004 mutation was observed with pGP202 and pGP204, whereas both pGP202 and pGP203 rescued the *S'am*2101 mutation. Thus, the *Sam*1004 mutation was mapped in the leftmost 600 bp and the *S'am*2101 mutation in the rightmost 900 bp of the G region. We identified in minicells the amber fragments of *Sam*1004 and *S'am*2101 which represent the translation products of the *S* and *S'* genes up to the amber mutations (Fig. 2). The amber fragment of *Sam*1004 has a molecular weight of 40,000, which requires a coding capacity of 1,100 bp of DNA. Thus, the site of the *Sam*1004 mutation is 1,100 bp removed from the translation initiation point of the *S* gene. The N-terminal end of the *S* gene should therefore be located at least 500 bp to the left of the G region. As 1,500 bp are needed to code for the *S* gene product (56K) this result agrees with the conclusion that the C-terminal end of *S* is located in the vicinity of the *KpnI* site in the G region. Taking into account the size of the *S'am*2101 amber fragment (45K, Fig. 2), we calculated in a similar way that the *S'* gene should start at least 330 bp to the left of the G region. From the results obtained so far, we conclude that the *S* and *S'* genes share a common part of DNA located outside G in the α region.

To map the promoter proximal end of *S* and *S'* more precisely we identified gene products and truncated proteins synthesized in minicells containing pGP56 and its derivatives containing Tn*5* in *S* or *S'* (Fig. 4). If truncated proteins are found, these will represent mainly truncated *S* or *S'* encoded amino acids, for it seems from the DNA sequence of Tn*5* (ref. 7) that in all three possible reading frames, translation stop codons occur within the first 30 bp from the ends of the transposon. As is shown in Fig. 4, lane 8, minicells containing pGP56 synthesize the 56K S polypeptide and the 50K S' polypeptide. Unfortunately, Tn*5* encodes polypeptides of reported[8] M_rs 54,000 and 49,000 which co-migrate with S and S' (Fig. 4, lanes 1, 2). Therefore, it was not possible to show the disappearance of S

Table 1 Mapping of the promoter proximal end of the *S* and *S'* genes

Plasmid	Position Tn*5* insertion (bp)*	M_r ($\times 10^3$) of truncated proteins†	Distance (bp) from the translation start point to inverted repeat‡ *S*	*S'*
pGP56::Tn*5* no. 33	400 from left IR	37	600	
pGP56::Tn*5* no. 3	575 from left IR	43	600	
pGP56::Tn*5* no. 63	875 from left IR	54	600	
pGP56::Tn*5* no. 69	825 from right IR	48		475
pGP56::Tn*5* no. 43	150 from right IR	25		525

* See Fig. 4*a* legend. IR, inverted repeat.
† Truncated proteins are indicated → in Fig. 4.
‡ Assuming an average molecular weight of 110 for an amino acid. The calculated distance is a maximal value as a contribution of translation into Tn*5* has not been taken into account.

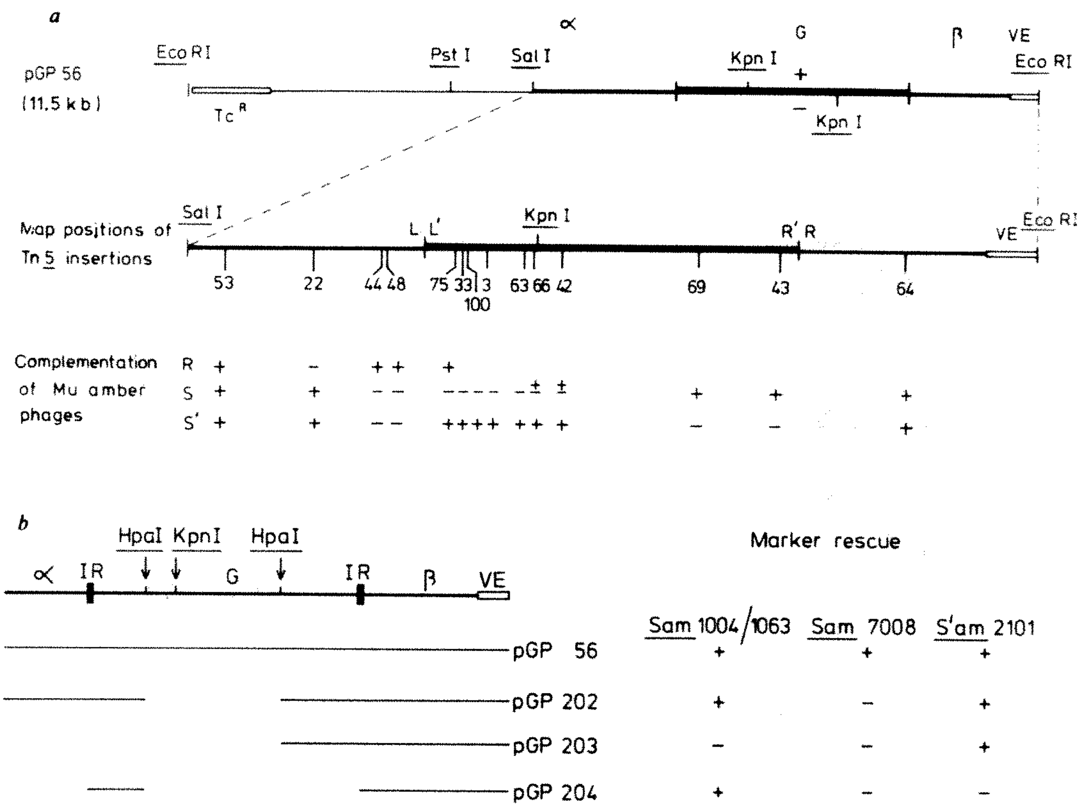


Fig. 3 *a*, Linear map of plasmid pGP56 (11.5 kb) which consists of Mu DNA (—) cloned in a *Pst*I–*Eco*RI fragment of pBR322(—). The Mu DNA comprises a part of α (3 kb), the G segment (3 kb), the β part (1.5 kb) and 0.4 kb of DNA from the variable ends (VE) of Mu DNA. A comparable plasmid has been described elsewhere[14]. The two possible orientations of the G segment in pGP56 preparations can be visualized by analysis of *Eco*RI/*Kpn*I fragments of pGP56 DNA. The map position of various insertions of transposon Tn*5* isolated in the *Sal*I/*Eco*RI fragment of pGP56 are indicated in the second line. The solid bar represents the G region and LL′ and RR′ the left and right inverted repeats. As restriction sites on Tn*5* have been presented[15], the integration sites of Tn*5* could be determined by digestion with restriction enzymes *Sal*I, *Eco*RI/*Xho*I, *Kpn*I and *Kpn*I/*Xho*I and length measurements of the obtained fragments on 1 and 1.4% agarose gels. As G inversion was observed within all plasmid preparations *Sal*I (which cuts once in Tn*5*), digestion shows whether Tn*5* is integrated in the G region. With *Kpn*I/*Eco*RI digestion, Tn*5* could be mapped to the left or to the right of the *Kpn*I site in G. The distance from the site of integration to the inverted repeats could be calculated by *Eco*RI/*Xho*I and *Kpn*I/*Xho*I digestion, as *Xho* sites in Tn*5* are known precisely[7] and the *Kpn*I site was mapped 975 bp from LL′. To measure complementation, *E. coli* K-12, *recA* or *E. coli* C, *recA* with and without pGP56 or pGP56::Tn*5*, were infected with Mu G(+) amber mutants (*Ram*1059 and *Sam*1004) or Mu G(−) amber mutants (*S′am*2101), respectively, with a MOI of 5. After an adsorption period of 15 min at 37 °C, the cells were incubated in the presence of Mu antiserum for 10 min at 37 °C, washed twice and diluted 1,000-fold in L broth. Phage titres were determined after 1 h growth at 37 °C by titration of phage on lawns of *E. coli* K-12, su$^+$ (for G(+) phage), or *E. coli* C, su$^+$ (for G(−) phage). The level of complementation is indicated + when the phage titres obtained are as high as that of pGP56 (5×10^8–1×10^9) and indicated (−) when they are as low as that of bacteria containing no plasmid (5×10^4–2×10^5). An aberrant level of Mu *Sam* complementation was obtained with Tn*5* at positions 66 and 42 (1×10^7 and 3×10^7, respectively). The reason for this low level of complementation is not known. *b*, Subclones of pGP56. The Mu DNA present on the different plasmids is indicated by a line. As pGP204 is derived from a G(−) molecule of pGP56 the β end is in fact directly linked to the left end of the G region (which is drawn here in the (+) orientation). The construction of the plasmids will be presented elsewhere (R.H.A.P. and P.P., in preparation). To measure marker rescue, *E. coli* K-12, su$^-$, or *E. coli* C, su$^-$, with or without a plasmid, were infected with Mu amber mutants with a MOI of 5. Titres of wild-type phage were determined by titration on lawns of *E. coli* K-12, su$^-$, or *E. coli* C, su$^-$. The *S* amber mutants used fall in different deletion groups[16].

or S′ when Tn*5* was inserted in the respective genes. However, truncated proteins are synthesized in minicells containing pGP56 with Tn*5* inserted at five different positions (Fig. 4, lanes 3, 4, 5, 7, 8). The C-terminal end of the *S* and *S′* genes could be mapped, respectively, less than 100 bp removed from Tn*5* no. 63 and Tn*5* no. 69, as their respective truncated proteins are only slightly smaller than the *S* and *S′* gene products (Fig. 4, lanes 5, 8). The N-terminal end, or translation startpoint, of *S* and *S′* could be calculated from the M_r of the various truncated proteins and the map positions of the Tn*5* insertions (Table 1). The position of the N-terminus of the *S* gene, ~600 bp removed from the G region, was also determined with pGP202 (Fig. 3*b*). Minicells containing pGP202 do not synthesize S or S′, confirming complementation data, but a 43,000 M_r polypeptide (Fig. 4, lane 10). Assuming that this polypeptide represents a truncated S protein, the *S* gene should start within 575 bp to the left of the inverted repeat. The results confirm that the start of both *S* and *S′* genes is located in the α region 500–600 bp away from the G region. Therefore, *S* and *S′* arise by a splicing of a common and an invertible part of DNA, as suggested in model B (Fig. 1).

With regard to *U* and *U′* in model B, we have previously been unable to prove genetically the existence of *U′* and to elucidate the expression of *U* and *U′* and their role in host specificity. However, in addition to S and S′, pGP56 encodes at least two polypeptides of M_r 22,000 and 18,000, which are not encoded by pGP202 (Fig. 4, lanes 9, 10). These polypeptides are good candidates for the gene products of *U* and *U′*. The 18K polypeptide may be the *U* gene product as it is not synthesized in minicells containing pGP56, Tn*5* no. 42 (Fig. 4, lane 6). Genetic experiments failed to prove that Tn*5* no. 42 is indeed integrated in the *U* gene, due to the leakiness of *Uam* mutants, although its physical position suggests this.

We have postulated that the *gin* gene is expressed, at least in part, from a promoter located in α (ref. 1). However, G inversion is observed with all derivatives of pGP56 described above. Assuming that the observed inversion is caused by *gin*[5,6], this result suggests that the *gin* gene is predominantly or entirely expressed from a promoter located to the right of Tn*5* no. 43, confirming the data of Kwoh *et al.*[9].

We believe the system described here to be the first example of gene splicing by an inversion event and the first example of

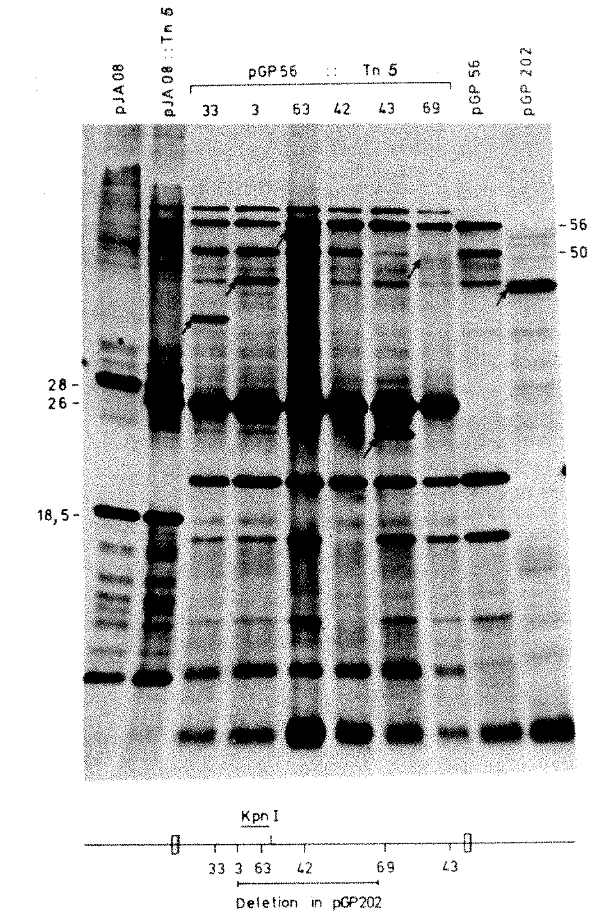


Fig. 4 Autoradiograph showing polypeptides synthesized in minicells isolated from cells containing pGP56, derivatives of pGP56 and plasmids with Tn*5* insertions. Minicells were labelled with ^{35}S-L-methionine and the polypeptides separated on a 12.5% polyacrylamide gel[13]. The autoradiograph shows the polypeptide pattern of minicells containing the following plasmids: control plasmid pJA08 (provided by R. Brandsma) (lane 1), pJA08::Tn*5* (lane 2), pGP56::Tn*5* no. 33 (lane 3), pGP56::Tn*5* no. 3 (lane 4), pGP56::Tn*5* no. 63 (lane 5), pGP56::Tn*5* no. 42 (lane 6), pGP56::Tn*5* no. 43 (lane 7), pGP56::Tn*5* no. 69 (lane 8), pGP56 (lane 9), pGP202 (lane 10). The *S* gene product (56K), *S'* gene product (50K), β-lactamase (28K)[13], neomycin phosphotransferase II (26K)[8] and the product of SSB (18.5K)[17] are indicated. Truncated proteins are indicated by →.

splicing within the prokaryotic cell. The recombination of a common region to different regions to form variable functional genes is a well known phenomenon in eukaryotes in the assembly of immunoglobulin genes[10,11]. In the case of G inversion, the variability is in the Sv or the Sv' region. The common region of the S and S' proteins is probably the part which becomes attached to the phage tail, whereas the variable part recognizes the receptor sites of the different hosts. This is now being tested by isolating and mapping mutants of Mu with other host ranges. Although we have shown that *S* and *S'* share a common region of DNA in α, this does not directly imply that their respective products are identical at the N-terminus. The calculated N-terminus of genes *S* and *S'* differed by ~75–100 bp. This may be due to errors in restriction mapping and/or molecular weight estimation. Moreover, we do not know exactly at which position translation has stopped in the various Tn*5* insertions. However, the observed difference may reflect a real difference in length of the Sc part of the two proteins. Therefore, definite proof for the presence of a common part in *S* and *S'* at the protein level must await determination of their amino acid sequences.

These results were first presented at the EMBO workshop 'Bacteriophage Mu' (May 1981) and were supported by the nucleotide sequence of the G region as presented by R. Kahmann (Max-Planck-Institute, München) at this workshop. We thank Ms J. Borst, W. Buys, S. Cramer, Ms A. Kuyper and H. Vrieling for their contribution to this work, Dr M. Howe for phage strains and Mrs N. van Hoek for typing the manuscript.

Received 14 September 1981; accepted 17 March 1982.

1. Van de Putte, P., Cramer, S. & Giphart-Gassler, M. *Nature* **286,** 218–222 (1980).
2. Howe, M. M. *Cell* **21,** 605–606 (1980).
3. Symonds, N. & Coelho, A. *Nature* **271,** 573–574 (1978).
4. Bukhari, A. I. & Ambrosio, L. *Nature* **271,** 575–577 (1978).
5. Kamp, D., Kahmann, R., Zipser, D., Broker, T. R. & Chow, L. T. *Nature* **271,** 577–580 (1978).
6. Chow, L. T., Kahmann, R. & Kamp, D. *J. molec. Biol.* **113,** 591–609 (1977).
7. Auerswald, E. A., Ludwig, G. & Schaller, H. *Cold Spring Harb. Symp. quant. Biol.* **45,** 107–113 (1981).
8. Rothstein, S. J., Jorgensen, R. A., Postle, K. & Reznikoff, W. S. *Cell* **19,** 795–805 (1980).
9. Kwoh, D. Y. & Zipser, D. *Virology* **114,** 291–296 (1981).
10. Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. & Tonegawa, S. *Nature* **286,** 676–683 (1980).
11. Davis, M. M. *et al. Nature* **283,** 733–739 (1980).
12. Howe, M. M., Schumm, J. W. & Taylor, A. L. *Virology* **92,** 108–124 (1979).
13. Giphart-Gassler, M., Wijffelman, C. & Reeve, J. *J. molec. Biol.* **145,** 139–163 (1981).
14. Schumann, W. & Bade, E. G. *Molec. gen. Genet.* **169,** 97–105 (1979).
15. Jorgensen, R. A., Rothstein, S. & Reznikoff, W. S. *Molec. gen. Genet.* **177,** 65–72 (1979).
16. O'Day, K., Schultz, D., Ericsen, W., Rawluk, L. & Howe, M. M. *Virology* **93,** 320–328 (1979).
17. Sancar, A., Williams, K. R., Chase, J. W. & Rupp, W. D. *Proc. natn. Acad. Sci. U.S.A.* **78,** 4274–4278 (1981).

Erratum

In the letter by P. M. Conn *et al., Nature* **296,** 653–655 (1982), the title should read 'Conversion of a gonadotropin-releasing hormone antagonist to an agonist: implication for a receptor microaggregate as the functional unit for signal transduction'. In Fig. 2 the lines labelled *a*, *b* and *c* should be termed (1), (2) and (3) to correspond to the text. In ref. 4 the page numbers should read 264–265, and the journal in ref. 6 is *Endocrinology* **109,** 2040–2045 (1981). In addition the following 'Note added in proof' should have appeared: Additional evidence for the significance of microaggregation comes from the observation of potency enhancement of a GnRH agonist in conditions which favour receptor microaggregation[20]. In addition, a hormone antagonist has been converted to an agonist by antibody cross-linking[21].

20. Conn, P. M., Rogers, D. C. & McNeil, R. *Endocrinology* (in the press).
21. Hopkins, C. R., Semhoff, S. & Gregory, H. *Phil. Trans. R. Soc.* B**296,** 73–81 (1981).

Corrigendum

In the letter 'Conversion of the chemical energy of methylmalonyl-CoA decarboxylation into a Na^+ gradient' by W. Hilpert and P. Dimroth, *Nature* **296,** 584–585 (1982), in Fig. 1 the total incubation volume should be 0.67 ml, not 10.67 ml as shown.

BOOK REVIEWS

Sociobiological sins

Michael Ruse

THESE two volumes, which should be read as one, are reports from a conference held in 1980 in Bresanone, Italy. As always with such reports, the quality of contributions is somewhat uneven and not everyone keeps quite so closely to the main themes as they should, in addition they are marked — and marred — by the leaden seriousness of the left. Nonetheless no one is left in any doubt as to the main motivating force behind the books, for, like King Charles's head, in virtually every contribution one topic keeps cropping up, namely the sins of sociobiology In response to this dreadful phenomenon, what the Dialectics of Biology Group offer us is one volume (*Against Biological Determinism*) directed to criticism of the suppositions behind sociobiology Then, the second volume (*Towards a Liberatory Biology*) turns to possibilities of something better

The critical volume wastes no time in setting the general theme, with an attack on human sociobiology as socially dangerous What worries Martin Barker is that sociobiology contains a "potential invitation to a unity of theory and practice in the latest forms of capitalist political economy" (p 26) This is a theme which we have heard sung many times before, most particularly by the Boston-based "Science for the People Sociobiology Study Group" I shall return to it later

Following more of the same, as well as a brief taxonomy by Hilary and Steven Rose of conventional attitudes to science, we arrive at what I find to be two of the more worthwhile papers, Lynda Birke's and Lesley Rogers's. These two feminists focus on attempts to explain gender identity and sexual orientation in terms of underlying biochemical causes (which in turn are supposedly linked to the genes) The well-known sex researcher, John Money, who must be rueing the day when E O Wilson spoke of his work favourably, is singled out for heavy criticism It is argued that Money bases his claims (specifically, that homosexuality has hormonal causes) on conceptual confusions, as well as inadequate evidence

The second charge cannot be denied Most sex researchers put me in mind of philosophers, as they generalize to the whole world from about two cases (which is one more case than Freud, who generalized exclusively from himself). But what of the more interesting claim of conceptual confusion? I am not sure that it holds. Is it really a mistake in logic to suggest that lesbians have "male" type emotions, or conversely that male homosexuals have female urges? What really counts in such a situation is the desire and the sex object towards which the desire is directed The actual mechanics involved seem to be irrelevant more an incidental by-product of physiology

Against Biological Determinism. Pp.184 Hbk ISBN 0-85031-423-2; pbk ISBN 0-85031-424-0 *Towards A Liberatory Biology* Pp 161 Hbk ISBN 0-85031-425-9; pbk ISBN 0-85031-426-7 The Dialetics of Biology Group, general editor Steven Rose (Allison and Busby/Schocken 1982) Each volume hbk £8 95, $14 95, pbk £4 50, $8 95

Moving along through slight variants on the anti-determinism (chiefly anti-sociobiology) theme in the first volume, we can turn to the more positive *Towards a Liberatory Biology.* Here we find what is perhaps the most sensible essay of the whole set "Synthesizing Views about the Origin of Behaviour" by Patrick Bateson The author perceptively points to some of the problems facing the student of behaviour, particularly those arising from the need to sort out innate from environmental factors. I suspect that Bateson must have felt somewhat uncomfortable at the Bresanone conference, for he calmly assumes that orthodox biologists have indeed made important advances in our understanding of behaviour. For instance, he praises Maynard Smith's use of game theory as a tool for understanding aggression, although elsewhere this is criticized as yet one more excursion into the realm of the unfalsifiable

Since sociobiology, with its emphasis on natural selection, is so very much part of the whole Darwinian paradigm, the most radical attack is obviously one which goes right to the heart of Darwinism itself, hoping to put forward a new theory of evolution This is the theme of several of the contributors to the second volume, as they urge upon us various neo-Lamarckian and neo-vitalist alternatives But I doubt that the Ernst Mayrs of this world need worry too much yet The critics base their case on some fairly worn arguments against natural selection, together with some incredible rewritings of history. I simply do not know how Gerry Webster and Brian Goodwin ("History and Structure in Biology") can accuse Darwin of being hostile to Cuvier's ideas or to the rational morphologists' "plan of creation" In fact, I would say that, in his thorough-going teleology, Darwin owes more to the non-evolutionist Cuvier than he does to the evolutionist Lamarck. Moreover, the very crux of Darwin's work was his incredible sensitivity to the morphologists' ideas, which he learnt from people like Richard Owen. Darwin's genius was to combine both form and function into one synthetic theory — form through common descent and function through natural selection Furthermore, to say that Darwin was "firmly in the empiricist/positivist tradition" (p 104) is to ignore entirely the influence on Darwin of such rationalist thinkers as William Whewell, and the difficulties that full-blown empiricists such as Huxley had with Darwin's selection mechanisms

All in all, I am afraid that I find these volumes unprepossessing; but let me make three concluding points First, I agree entirely with the DBG that a blind reductionism, aiming to explain everything in terms of the Hardy-Weinberg law — or even worse, some simplistic law of physics — is the last thing that biologists should aim for Anyone who says that life and its problems is just a set of molecules is either a fool or a liar But, second, there is no evidence that sociobiology is committed to such a gormless programme of reductionism, despite the outlandish statements of some of its more enthusiastic proponents Sociobiologists continue to deal with *biological* questions, moreover, the last thing that sociobiology does is to exclude methods and findings from other disciplines, such as psychology and sociology Rather, the social sciences are given a crucial role in the overall picture Thus, for instance, suppose that there is something in the sociobiological claim that incest barriers have a genetic foundation This says nothing about the role which genotypes have in causing such barriers Instead, all sorts of questions are asked (and answers demanded) about the ways in which close contact with siblings in early childhood lead to anti-sexual imprinting And who is to answer these questions but social scientists?

This conclusion is precisely what one would expect, given that sociobiology lies so firmly within the Darwinian tradition Darwinism has always claimed that evolution is a function of heredity in interaction with the environment. A purely biological (meaning especially a purely genetic) determinism is meaningless to the Darwinian

Third and finally, let me speak to the concern articulated by Martin Barker, but I think behind nearly all of the opposition to sociobiology and like enterprises When all is said and done, is it not really the case that sociobiology is merely right-wing ideology dressed up to look like genuine science? One would have to be insensitive not to recognize this as a genuine concern, or to deny that genetic theories of human behaviour have been (and still are) used to "support" some pretty horrible social doctrines But it is simply not true to claim unequivocally that all genetic theories are apologia for the worst excesses of Western capitalism, any more than it would be true to claim that all environmentalist theories are as ideologically pure as the driven snow Indeed, some thoroughly dreadful practices have been pushed by people who claim that all we need is a little of the right social conditioning

In short, I remain unconvinced that sociobiology is the inherently evil thing supposed by the DBG No doubt the members of the DBG will remain equally unconvinced of what I say But could I make one request? Do not let us degenerate into accusations that the other side is responsible for the rise and success of Creationism — as the cover of one of these books suggests There is simply no causal connexion between sociobiology and Creationism, and to pretend otherwise is gratuitously insulting For all our differences, do not forget that we are all striving, painfully, to see the light □

Michael Ruse is a Professor in the Departments of History and Philosophy at the University of Guelph His most recent book is Darwinism Defended *(Addison-Wesley, 1982)*

Journals' review issue 1982

On October 7th *Nature* will publish a second review supplement devoted to science journals Last year's supplement, covering journals first published between January 1978 and May 1980, appeared in *Nature* 293, 341-369, see p 343 of that issue for details

Criteria for inclusion of a journal in the 1982 issue are that

- the first number appeared, or the journal was re-titled, between June 1980 and May 1981,
- it is published at least three times a year,
- the main language used is English

Broadly, periodicals of professional interest to scientists will be considered for review, with the exception of abstracts' journals

In addition it is hoped to cover publicly available newsletters, first published between January 1978 and May 1981, in that issue

Publishers and societies are invited to submit four sample issues of periodicals satisfying the above criteria, *including the first and most recent numbers*, to the Review Editor, *Nature*, 4 Little Essex St, London WC2R 3LF, England (London 836 6633 ext 2570) as soon as possible

The biological roots of personality

Anthony Gale

A Model for Personality. Edited by H. J Eysenck Pp 287 ISBN 3-540-10318-X/ 0-387-10318-X (Springer-Verlag 1981) DM78, $37 20

AT THE dinner last summer to celebrate the publication of his *Festschrift*, Hans Eysenck declared himself to be confused This unusual state of mind resulted from a paradox The same scientific thinking which had led him to reject fascism in his youth had created enemies in his old age His analysis of IQ differences between ethnic groups, and his insistence that biological factors account for differences between individuals, have led to accusations of racism Eysenck-bashing has become a popular sport, not only on the campus, but in scientific journals He only needed to look to his own writings for an answer to the paradox, frequently he talks of the *zeitgeist*, the prevailing scientific fashion which favours some approaches and castigates others And he has never been afraid to profess unpopular views, whether it be to attack psychoanalysis or to accuse his fellow psychologists of sloppy thinking

In *A Model for Personality* Eysenck invites his critics to do their worst His theory of the biological basis of extraversion, and the data which support it, are laid out before us, warts and all He sees himself in a long line of biological theorists, stretching back to the Greek physician, Galen The argument is straightforward systematic studies of personality reveal two common factors, extraversion–introversion and neuroticism–stability, and their ubiquitous appearance implies that they are *given* by nature If they have biological roots, we expect them to be normally distributed across the population, to be transmitted by genetic mechanisms, expressed in brain function, and adaptive in governing the interaction of the individual with his physical and social environment Each of the contributors to the book (all are internationally established authorities) deals with one of these aspects of the theory

The reader should recognize that, for many psychologists, the establishment of links between genetics, biology, learning, and the way people think and interact socially involves a tremendous imaginative leap into the unknown, for it presupposes that all the philosophical problems raised by the Mind–Body relationship will ultimately be resolved Some of the tricks, which involve a shift between diverse domains of description, are truly remarkable For example, a 24-item *paper and pencil* test about how you feel in different situations ("Would you rather stay at home, or go to a dull party?") enables successful prediction of *electrical* properties of the brain Eysenck believes, as did Pavlov, that nervous systems *qua* nervous systems, have properties which determine the ways in which the organism acts upon and responds to the environment *The Dynamics of Anxiety and Hysteria* (Routledge & Kegan Paul, 1957) and *The Biological Basis of Personality* (Thomas, 1967) showed how Eysenck's theory could find operational expression in the form of systematic laboratory experiments This *rapprochement* between traditional experimental psychology and personality theory is illustrated by his son Michael's chapter, reviewing studies of personality, learning and memory

The penultimate chapter, by Jeffrey Gray, offers a savage critique of the theory He claims that Eysenck's selection of a particular mathematical solution to factor data does not fit the facts, that every data set supporting the theory is matched by a contradictory set, that the 1957 and 1967 versions generate different and even conflicting predictions, and that the key instrument of measurement (the Eysenck Personality Inventory) has undergone surreptitious changes over the years Only Eysenck could invite a colleague (possibly his successor at the Institute of Psychiatry?) to write such a provocative and thoroughgoing critique of his own theory

This is a very useful, balanced and sensibly compiled record of Eysenck's progress in this his retirement year It is ironic that in the 1980s there is a move towards a more biological perspective the pendulum is swinging back in Eysenck's direction □

Anthony Gale is Professor of Psychology at the University of Southampton He is co-editor of The Biological Bases of Personality and Behaviour *(Hemisphere Press, 1982) and* Psychology and People A Tutorial Text *(BPS–Macmillan, 1982)*

Chariots of water

M. Ian Phillips

Thirst By Barbara J Rolls and Edmund T Rolls Pp 194 Hbk ISBN 0-521-22918-9, pbk ISBN 0-521-29718-4 (Cambridge University Press 1982) Hbk £15, $29 50, pbk £5 50, $10 95

EVERY year in London on the neutral waters of the Thames there is a race between a row-boat from Cambridge University and one from Oxford I was reminded of this when asked to review *Thirst* by Barbara Rolls and Edmund Rolls from Oxford University, since a year ago I

reviewed *The Physiology of Thirst and Sodium Appetite* by James Fitzsimons from Cambridge University (*Science* **208**, 711, 1980)

This latest response from Oxford in the race for dominance of the water intake field is a fast-moving, sometimes too brief but nonetheless highly readable account of our present knowledge of thirst. It is not as erudite as the volume by Fitzsimons — it is less than half the size for one thing — but it is a book that one could recommend to students of physiological psychology and researchers newly attracted to the field It covers a comprehensive list of topics and as an introductory text it is excellent What it lacks for the researcher actively involved in these studies is the depth of sources and an accounting of all available information For example, the authors avoid covering the brain angiotensin system in detail by dismissing it as being controversial Perhaps that attitude is the better part of valour, but in consequence much recent data are overlooked

Since most of the interest in the subject can be attributed to the discovery that injections of angiotensin II stimulate thirst, it is appropriate that Barbara Rolls is the senior author of this text It was she, with Fitzsimons, who first showed that effect of angiotensin II intravenously and shortly afterwards intracranially Edmund Rolls adds his unique expertise in the understanding of central mechanisms of thirst in primates It is a winning combination for a work that will appeal to those interested in the physiology of motivation

The book is profusely illustrated with clear pictures, almost one to a page. The material presented is a readable summary of the progress that has been made in the study of thirst and particularly since the discovery of the effects of angiotensin II in 1969 It is apparent, however, that the race is not over because the control of drinking under normal physiological conditions has not yet been established, the authors say that involvement of angiotensin in thirst may only be in emergencies As a review of the field, the book enables one to see that advances have been made with relatively simple techniques The flow of experiments presently is towards more precise localization of receptors and more accurate descriptions of the brain pathways involved. In that respect, thirst research is not yet far from the starting line

The book does show that there is sufficient data for a course in thirst and for an active field of research There is certainly room for the two volumes on water intake without causing hypervolumia and, as far as competition between the two books goes, the winner seems to be Cambridge since its University Press is the publisher of both. □

M Ian Phillips is Chairman of the Department of Physiology at the University of Florida, Gainesville

Genes through the looking glass

P.H. Williams

Gene Function E coli *and Its Heritable Elements*. By Robert E. Glass Pp 487 Hbk ISBN 0-7099-0081-3; pbk ISBN 0-7099-0082-1 (Croom Helm/University of California Press 1982) Hbk £19 95, $45 60, pbk £9 95, $22 80

WHEN, several years ago, I embarked upon post-graduate research in microbial genetics, the pre-eminent textbook of the field was *The Genetics of Bacteria and their Viruses* by William Hayes To a newcomer like myself, Hayes was indispensable, offering all there was to know about molecular genetics in an authoritative, well-referenced, easily readable package It was with great regret that I watched that book drift slowly, but inevitably, out of date

Gene Function by Robert Glass is, to use Hayes's description of his own work, "a rather advanced text book" It covers much the same ground as Hayes, and is aimed at the same type of audience, namely at advanced undergraduates, and at researchers in the field at every level. Only a few authors since Hayes have successfully abridged the monumental mass of *Escherichia coli* research into a single informative but highly readable volume In my opinion Robert Glass has made a creditable attempt

Gene Function, however, is not merely Hayes updated Over and above the incorporation of the wealth of knowledge accumulated over a decade and a half, Glass's book has an entirely different viewpoint which reflects the backgrounds and attitudes of those who have more recently entered and contributed to the field Hayes began his book with an introduction to genetic principles, and then developed the theme of bacteria and bacteriophage as genetic systems Glass, on the other hand, writes as a biochemist, beginning with an overview of the structure and biological role of nucleic acids at the molecular level (or rather a preview, since these are covered in detail later), before outlining elements of microbial physiology and metabolism essential to a full comprehension of experimental methods

In the preface to his book the author thanks a colleague for "a 'sensible' title" It is a little surprising that a volume devoted exclusively to one organism should presume to claim for itself the title *Gene Function* However, this book is about genes and the way they function in that particularly convenient little bag known as *E. coli* — simple genes, perhaps, organized and regulated differently from those of eukaryotes, certainly, but the strong implication is that knowledge of the molecular wonderland of *E coli* genes will form a firm basis for the study of genes in any other organism Nonetheless, the title still worries me a little.

Gene Function is well researched, adequately referenced and nicely balanced in terms of breadth of material and depth of detail. In general it is clearly written, although extensive cross-referencing tends to destroy the reader's train of thought at times The frequent use of footnotes and parentheses in the text irritated me somewhat, and I confess that I found the index almost impossible to use However, these are quite minor quibbles, and, except for the last point, did not seriously detract from my enjoyment of the book. *Gene Function* may not, perhaps, be as fondly remembered 15 years hence as Hayes's book is now, but there is no doubt that it is an important and useful addition to the book-shelves of today's molecular geneticists I hope it can be regularly updated without losing any of the evident enthusiasm of its author. □

P H Williams is a Lecturer in the Department of Genetics at the University of Leicester

Space-time through time

Paul Davies

The Science of Space-Time By Derek J Raine and Michael Heller Pp 255 ISBN 0-912918-12-8 (Pachart Publishing, Tucson, Arizona. 1981.) $24.

IT IS curious the way that certain ideas and concepts in science exercise a wide and seemingly disproportionate fascination. These include the measurement paradoxes of quantum theory, and the origin of time's mysterious "arrow" Another is Mach's principle Scientists return to it again and again Now, for the first time, a comprehensive treatment of Mach's principle is available in a book that can be profitably read by those with only a basic knowledge of space-time physics

The basic idea of Mach's principle is that the property of inertia in a material body has its origin in cosmology A rotating planet, for example, feels centrifugal stresses and bulges at the equator But visually the Earth's rotation is deduced by "watching the stars go round" So, are these two radically different phenomena — the experience of centrifugal bulging and the whirling of the stars — in some way connected? Are rotations (and other types of accelerated motions) purely relative to the background framework of far-flung galaxies?

Einstein at first thought so, and Mach's work deeply influenced his development of the theory of relativity In 1916, on Mach's

death, Einstein wrote a tribute to him But in spite of this Machian input, a clear-cut manifestation of Mach's principle failed to emerge in any very obvious way from relativity theory Over the decades, many theoretical physicists have struggled to find formulations of Mach's principle that would mesh easily with either the theory of relativity, or some close variant It is a task as yet unfinished, but it remains as compelling as ever At issue is one of the fundamental properties of space-time structure, and the nature of motion

This timely and most welcome book contains a discussion of the evolution of the space-time concept from Aristotle to Einstein (and beyond), and presents a detailed and up-to-date analysis of Mach's principle in a modern setting The authors conclude that "At best, Mach's principle has been expressed in general relativity as a selection rule To go beyond this one may have to give up general relativity"

However, this book is far from a catalogue of failures It is a comprehensive and accurate survey of the physics of space and time and of the philosophical overtones which accompany it. The theory of relativity, both the special and general versions, obviously plays a central role in these ideas, and the reader will find discussions of black holes and details of gravitation experiments as well as a careful introduction to the more elementary aspects of space-time structure

Mach's principle is an intriguing context in which to learn relativity and space-time physics, as well as an end in itself I can thoroughly recommend this book to undergraduates learning general relativity, as well as to researchers who share the authors' fascination with one of science's age-old mysteries □

Paul Davies is Professor of Theoretical Physics at the University of Newcastle upon Tyne His most recent book (with N D Birrell) is Quantum Fields in Curved Space *(Cambridge University Press, 1982)*

The mixed blessing: oxygen in biology

Barry Halliwell

Oxygen and Living Processes An Interdisciplinary Approach Edited by Daniel L Gilbert Pp 401 ISBN 3-540-90554-5/0-387-90554-5 (Springer-Verlag.1981) DM135, $61 40

THIS colourless, odourless gas is an extreme fire hazard and is very toxic to some organisms at minute concentrations Its toxicity to human beings is much slower-acting, but nevertheless over a period of many years it produces a progressive oxidative deterioration that results inevitably in death Only the swim-bladders of certain fishes can tolerate it at high concentrations, although turtles and crocodiles are fairly resistant to its toxic effects unless they are placed in warm water

The above description, using information taken from this book, applies of course, to oxygen Oxygen is essential for efficient energy production by oxidation of foodstuffs in aerobic animals, but only its peculiar chemistry protects us from instant "spontaneous" combustion The same chemistry dictates that the oxygen reduction accompanying substrate oxidation must proceed via partially reduced oxygen intermediates on the pathway to water Such intermediates as superoxide and hydroxyl radical are particularly noxious and so aerobic cells are equipped with many defences against them, including superoxide dismutase, catalase, glutathione and/or ascorbate peroxidases and vitamin E The cytochrome oxidase complex of mitochondria is carefully designed not to allow the escape of such intermediates from its active site

As its title declares, this book is indeed an interdisciplinary approach to oxygen All of the items I have mentioned are covered in detail, and in addition there are chapters on photosynthetic oxygen production, the geological history of the oxygen in the Earth's atmosphere, oxygen-carrying pigments, kinetics of oxygen diffusion, the history of the discovery of oxygen, and the medical uses (and abuses) of the gas, covering blindness in babies, pulmonary oxygen toxicity and hyperoxygenation in the treatment of gas gangrene and cancer of the head and neck The contributions are well-written, all of them by experts in specific fields, and the book has been produced in an attractive style with the minimum of errors

I do have some minor niggles, however Much of Chapter 1, a history of the discovery of oxygen, is merely an over-condensed — and rather boring — list of names and dates, for example, it is impossible to do justice to the complex physiology expounded by Galen in a few lines In Chapter 6, on photosynthesis, I doubt whether non-biologists will really understand what exactly are photosystem I, photosystem II and Q In addition, the order of the chapters is sometimes a bit illogical

These comments are not meant to detract from what is an excellent book It is truly "everything you wanted to know about oxygen, but didn't know whom to ask" □

Barry Halliwell is a Lecturer in Biochemistry at King's College, University of London, and author of Chloroplast Metabolism *(Clarendon/Oxford University Press, 1981)*

ANNOUNCEMENTS

Awards

The French National Order of the Legion of Honor has been awarded to University of California, San Francisco clinical professor **Piero Mustacchi** for 25 years of prominence in professional activity, in particular for developing academic and medical links between the University of California and universities in France and French-speaking Africa

Professor William H. Thorpe (University of Cambridge) has been presented with the annual prize of the Fyssen Foundation for his research in animal behaviour

Dr Prescott L. Deininger (Louisiana State University Medical Center in New Orleans) and **Dr Jeffrey W. Williams** (Ohio State University) have been selected to receive the Boehringer Mannheim Biochemicals travel awards in molecular biology and in enzymology

The British Society of Rheology has awarded its annual award to **Professor Joachim Meissner, Dr Hans Martin Laun, Dr Helmut Munstedt** and **Dr Manfred Wagner** for work in the development of elongational rheometry

Five prizes have been awarded to young Wiezmann Institute scientists the Somach Sachs Memorial Award to **Dr Itamar Procaccia** for studies on rates of chemical reaction near phase transition points, the Sarah Zinder Leedy Memorial Award to **Dr Gideon Schechtman** for work on geometric structure of Banach spaces, the Samuel Yaroslavsky Memorial Award to **Dr Abraham Shanzer** for work on synthesis of large, ring-shaped organic materials, the E D Bergmann Award to **Dr Michael Levitt** for studies of conformation and structure of macro-molelcules, the M Levinson Prize to **Dr Mordechai Milgrom** for work on emission of spectral lines with changing wavelengths by the astronomical object SS 433

Twenty Harkness Fellowships tenable for between 12 and 21 months are offered each year The award includes return fares to the United States, living and family allowances, travel in America, tuition and research expenses, equipment allowance and health insurance Candidacy is open to men and women in any profession or field of study, who are citizens of the UK and have received both secondary schooling and further education (or equivalent professional experience in lieu of further education) wholly or mainly in the UK Candidates must be between 21 and 30 years of age on 1 September 1983, unless qualified in medicine or employed in the Civil Service or the media, in which cases the upper age limit is 33 For application forms (available only to candidates) send a self-addressed envelope carrying 29p postage, and measuring not less than 10 by 7 inches, to The Harkness Fellowships (UK), Harkness House, 38 Upper Brook Street, London W1Y 1PE Forms will not be posted from the office after 7 October, 1982

In recognition of his contributions to the understanding of the meaning of evolution, the history of life, animal taxonomy and mammalian systematics, The Peabody Museum of Natural History, Yale University has established the George Gaylord Simpson Prize in celebration of his 80th birthday, 16 June 1982 The prize will be presented annually by the Peabody Museum to a young author (under 35) of an outstanding paper in the above areas of science Applications to the Director, Peabody Museum of Natural History c/o Yale University, PO Box 6666, 170 Whitney Ave, New Haven, Connecticutt 06511, USA by 31 December 1982

Appointments

The National Academy of Sciences has announced the election of the following foreign associates **Nicola Cabibbo** (theoretical physics, University of Rome), **Shmuel Eisenstadt** (sociology, Hebrew University of Jerusalem), **Paul Fraisse** (experimental psychology, University of Paris V), **Marianne Grunberg-Manago** (CNRS, Paris, and biochemistry, University of Paris VII), **Hua Luogeng** (mathematics, Chinese Academy of Sciences), **Il'ya Mikhailovich Lifshitz** (theoretical department, Institute for Physical Problems, Moscow), **Jacques F.A.P. Miller** (experimental pathology, The Walter and Eliza Hall Institute of Medical Research, Melbourne); **Martin J. Rees,** (theoretical astronomy, University of Cambridge), **Ralph Riley,** (Agricultural Research Council of the UK); **John Maynard Smith** (biology, University of Sussex), **Takashi Sugimura** (molecular biology, Tokyo University); **Tsuneo Tomita** medicine, Keio University, Tokyo)

Professor Emanuel Mazor, has been named as the first incumbent of the Frank W Considine Chair in Hydrological Research at the Weizmann Institute of Science, Rehovot

Paul C. Martin will occupy the newly established chair at Harvard University the John Hasbrouck Van Vleck Professorship of Pure and Applied Physics

Dr William Taylor present director of the University of London Institute of Education has been appointed principal of the University of London from 1 September 1983 in succession to Mr J R Stewart who will be retiring

Meetings

21-24 June, **Environment 82,** Stockholm (Environment 82, Ministry of Agriculture, S-103 33 Stockholm, Sweden)
28 June-2 July, **Meeting on The Monsoon Climate Programme,** Geneva (WMO Secretariat, Geneva, Switzerland)
28-30 June, **Nordic Congress of Pathological Anatomy and Cytology,** Copenhagen (NOPAC '82 Secretariat, Institutterne, Frederik den V's vej 11, DK-2100 Copenhagen O, Denmark)
29 June, **Design and Safety of the Sizewell Reactor,** London (The Royal Society, 6 Carlton House Terrace, London SW1, UK)
29-30 June, **Food and People,** London (The Secretary BNF, 15 Belgrave Square, London SW1, UK)
30 June, **Energy Conservation, Economic Growth, and Employment,** London (Centre for Energy Studies, Polytechnic of the South Bank, Borough Rd, SE1, UK)
30 June-1 July, **Trace Elements in Animal Production and Vetinary Practice,** York (BVA, 7 Mansfield St, London W1, UK)
28 June-2 July, **Biomedical Writing,** Harvard (Office of Continuing Education, Harvard School of Public Health, 677 Huntington Ave, Boston, Massachusetts 02115, USA)
5-7 July, **Distributed Computing Systems,** Strathclyde (Mr F Chambers, Logica, 64 Newman St, London W1, UK)
5-9 July, **Concrete Mix Design,** London (Cement and Concrete Association, Fulmer Grange, Fulmer, Slough, UK)
6-9 July, **North Atlantic Ocean Stations,** Geneva (World Weather Watch Dept, WMO, Geneva, Switzerland)
8-9 July, **Physical Techniques in Cardiological Imaging,** Southampton (Dr M D Short, Dept of Medical Physics, University College Hospital, London WC1, UK)
9-10 July, **Genetics Cancer–Environment,** Monaco (Institut Pasteur de Lyon, 77 rue Pasteur, 69365 Lyon Cedex 2, France)
12-16 July, **Environment and Safety Risk Management,** Massachusetts (Director of the Summer Session, Room E19-356, MIT, Cambridge, Massachusetts 02139, USA)
13-16 July, **Proteins and Nucleic Acids in Plant Systematics,** Bayreuth (Prof Dr U Jensen, Dept of Plant Ecology and Systematics, University of Bayreuth, PO Box 3008-D-8580 Bayreuth, FRG)
14-15 July, **Quantatitive Structure Activity Relationships in Taste and Olfaction,** Reading (Dr M G Lindley, Tate & Lyle Group R & D, PO Box 68, Whiteknights, Reading, UK)
14-16 July, **Ions, Cell Proliferation and Cancer,** Lake Placid (Dr W L McKeehan, W A Jones Cell Science Center, Old Barn Rd, Lake Placid, New York 12946, USA)

15-16 July, **Climatic Effects of Increasing Atmospheric Carbon Dioxide,** East Anglia (Dr T M L. Wigley, Climatic Research Unit, University of East Anglia, Norwich, Norfolk, UK)

19-23 July, **Basic Cell and Organ Culture Course,** Boston (Course Administrator, Cell Culture Research Unit, Dept of Anatomy and Cellular Biology, Tufts University School of Medicine, Boston, Massachusetts 02111, UK)

19-22 July, **Computer Aided Design and Manufacture,** Manchester (S Mayhew, Scientific and Technical Studies, Norwich House, 11/13 Norwich St, London EC4, UK)

20-22 July, **Fetal Antigens and Cancer,** London (The Ciba Foundation, 41 Portland Place, London W1, UK)

23-24 July, **Immune Complexes: Formation, Biological Properties and Clinical Aspects,** Nottingham (Dr W G. Reeves, Dept of Immunology, University Hospital, Nottingham, UK)

10 August, **Milestones and Trends in Polymer Science,** Michigan (Dr J K Rieke, Chemical Company, Midland, Michigan 48640, USA)

23-28 August, **Stress Effects on Photosynthesis,** Diepenbeck (Dr R Marcelle, Laboratory of Plant Physiology, Research Station of Gorsem, B-3800 Sint Truiden, Belgium)

23-27 August, **9th Annual European Geophysical Society Meeting,** Leeds (Royal Meteorological Society, James Glaisher House, Grenville Place, Bracknell, Berks, UK)

23-26 August, **Powder Mixing Technology,** Central New Jersey (The Center for Professional Advancement, PO Box H, East Brunswick, New Jersey 08816-0257, USA)

29 August-4 September, **21st International Horticultural Congress,** Hamburg (Hamburg Messe und Congress Gmbh Abt. Offentlichkeitsarbeit, Postfach 30 23 60, 2000 Hamburg 36, FRG)

20-24 September, **Molecular Crystal Symposium,** Quebec (D Chartrand, National Research Council Canada, Ottawa, Canada K1A 0R6)

21-23 September, **Radiation Curing,** Illinois (S Buhr, Technical Activities Dept, Society of Manufacturing Engineers, 1 SME Drive, PO Box 930, Dearborn, Michigan 48128, USA)

21-24 September, **Ceramic-Metal Joining,** Amsterdam (The Center for Professional Advancement, Postbus 19865, 1000 GW Amsterdam, The Netherlands)

21-24 September, **Pteridines and Folic Acid Derivatives,** St Andrews (Prof H C S Wood, Dept of Pure and Applied Chemistry, University of Strathclyde, 295 Cathedral St, Glasgow, UK)

21-24 September, **Wind Energy Systems,** Stockholm (BHRA Fluid Engineering, Cranfield, Bedford, UK)

21-25 September, **British Veterinary Association, Centenary Congress,** Reading (BVA, 7 Mansfield St, London W1, UK)

22-24 September, **601st Meeting of The Biochemical Society,** Aberdeen (The Meetings Officer, The Biochemical Society, 7 Warwick Court, London WC, UK)

22-26 September, **Biophysical Aspects of Receptor Structure and Function,** Portorose (Prof. D. Hadzi, Bopis Kidric Institute of Chemistry, POB 380, 61001 Ljubljana, Yugoslavia)

23 September, **Fluids in Metamorphism,** Glasgow (Dr M. Brown, Dept of Geology and Physical Sciences, Oxford Polytechnic, Oxon, UK)

23-24 September, **Choice of Material for Condenser Tubes and Plates and Tube and Tightness Testing,** Avignon (Societe Francaise D'Energie Nucleaire, 48 rue de la Procession, F-75724 Paris, Cedex 15, France)

27 September-1 October, **1st International Congress of Neuroimmunology,** Lake Maggiore (Dr P O Behan, Institute of Neurological Sciences, Southern General Hospital, 1345 Govan Rd, Glasgow, UK)

28-30 September, **The Geological Evolution of the Eastern Mediterranean,** Edinburgh (Dr A,H F Robertson, University of Edinburgh, Grant Institute of Geology, West Mains Rd, Edinburgh, UK)

27 September-1 October, **Physical Chemistry of Transmembrane Ion Motions,** Paris (Dr C Troyanowsky, Societe de Chimie Physique, 10, rue Vauquelin, F75321 Paris, Cedex 05, France)

28-29 September, **Particle-Fluid Separation in Offshore Engineering,** Middlesborough (Dr G S Bainbridge, Dept of Chemical Engineering, Teesside Polytechnic, Borough Rd, Middlesborough, Cleveland, UK)

29 September-1 October, **Biotechnology in Clinical Chemistry,** Cambridge (Dr C P. Price, Dept of Clinical Biochemistry, Addenbrooke's Hospital, Hills Rd, Cambridge, UK)

30 September-2 October, **Deburring and Surface Conditioning,** New Orleans (J Slaughter, SME, 1 SME Drive, PO Box 930, Dearborn, Michigan 48128, USA)

4-6 October, **Technology for Space Astrophysics — The Next 30 years,** Danbury (N Geril, Perkin-Elmer Corp, 100 Wooster Heights Rd, Danbury, Connecticut 06810, USA)

4-7 October, **Hormones and Cell Regulation,** Alsace (Dr R Denton, Dept of Biochemistry, University of Bristol, Medical School, Bristol, UK)

5-8 October, **Leaf Protein Research,** Aurangabad, (Dr N Singh, Central Food Technological Research Institute, Mysore 570 013, India)

6-8 October, **The Chemical Industry and the Environment in the 1980s,** London (Society of Chemical Industry, 14 Belgrave Square, London SW1, UK)

12 October, **Radiological Protection and Home Defence,** London (Prof J H. Martin, Dept of Medical Biophysics, Dundee University, Dundee, UK).

13-15 October, **Energy: Money, Materials and Engineering,** London (The Institution of Chemical Engineers, George E Davis Building, 165-171 Railway Terrace, Rugby, UK)

17-22 October, **Congress of Allergology and Clinical Immunology,** London (XIIACACI, Conference Associates ICACI, 34 Stanford Rd, London W8, UK)

18-20 October, **New Spectroscopic Methods for Biomedical Research,** Columbus, (K L. Waite, Batelle's Columbus Laboratories, 505 King Ave, Columbus, Ohio 43201, USA)

18-20 October, **Radar '82,** London (The Institution of Electrical Engineers, Savoy Place, London WC2, UK)

24-28 October, **14th Meeting of the European Tumor Virus Group,** Bulgaria (Dr T G Todorov, Bulgarian Academy of Sciences, Inst. of General and Comparative Pathology, Acad, G Bonchev str Blok Ill, Sofia 1113, Bulgaria).

26-28 October, **Medical Imaging and Image Interpretation,** Berlin (Dr M.H Loew, Dept of Electrical Engineering and Computer Science, George Washington University, Washington DC 20052, USA)

26-28 October, **Polynuclear Aromatic Hydrocarbons,** Columbus (Dr M Cooke, Battelle's Columbus Laboratories, 505 King Ave, Columbus, Ohio 43201, USA)

29-31 October, **3rd General Meeting of the American Pain Society,** Miami (K L Casey, Chief Neurology Service, VA Medical Center, 2215 Fuller Rd, Ann Arbor, Michigan 48105, USA)

31 October-2 November, **The Society for Environmental Geochemistry and Health, 1st Annual Scientific Meeting,** North Carolina (C Harrington, Dept of Surgery, School of Medicine, East Carolina University, Greenville, North Carolina 27834, USA)

1-5 November, **Effects of Ultraviolet Radiation on Plants,** Delhi (Prof L O Bjorn, Dept of Plant Physiology, Box 7007, S-220 07 Lund, Sweden)

7-10 November, **Woodlands Conference on Sustainable Societies,** Texas (The Woodlands Conference, PO Box 9663, Arlington, Virginia 22209, USA)

8-10 November, **Applications of Accelerators in Research and Industry,** Texas (J L Duggan, Dept of Physics, PO Box 5368, North Texas State University, Denton, Texas 76203, USA)

7-10 November, **1st International Workshop on Human Leucocyte Differentiation Antigens,** Paris (Dr A Bernard, Hôpital Saint-Loois, 2 Place du Docteur Fournier, 75475 Paris, Cedex 10, France)

16-19 November, **Applications of Biological Markers to Carcinogen Testing,** Bethesda (The Council for Research Planning in Biological Sciences, Inc c/o Associated Universities, Inc 1717 Massachusetts, NW 603, Washington DC 20036-2077, USA)

nature

3 June 1982

Squabbles over East–West pipeline

The Soviet plan to build a huge pipeline for natural gas for customers in Western Europe has irked the United States Administration. But Washington's fears are exaggerated.

The Western economic summit that opens tomorrow (4 June) at Versailles is likely to be more than usually contentious. President Ronald Reagan, one of the several first-time participants at these occasions, is unlikely to find the meeting congenial, if only because the United States will find itself hemmed firmly into a corner about its distinctive (some say eccentric) economic policy, a curious blend of Keynes and Friedman whose chief external effect is to keep international interest rates so high that economic recovery will be impeded. (The Prime Minister of Japan, Mr Zenko Suzuki, will, on the other hand, be pilloried for what he may fairly claim is a paradoxical reason — interest rates notwithstanding, Japan has been too successful.) But when the going gets rough, the United States will reply with its now long-standing complaint that Western Europe should not be preparing to buy natural gas from the Soviet Union. This issue, about which Washington has been sucking its teeth off and on for the best part of a year, is without doubt important. But not in the sense in which the United States appears to believe.

The issue exists because of the discovery of substantial reserves of natural gas in the Soviet Urengoi field, in the southern Urals. The plan is to ship large quantities of this gas westwards towards Europe (East and West) by means of a huge pipeline on which construction is about to start. West Germany, France and Switzerland have already signed contracts to purchase natural gas over the next thirty years or so. American objections to European involvement in this scheme take three forms. First, the argument goes, Western European dependence on natural gas from the Soviet Union will undermine its strategic independence. Second, the method by which the pipeline will be financed, with the eventual customers providing credits for the construction, being repaid with gas at a later stage, makes it possible for the Soviet Union to avoid what would otherwise be a nasty brush with the international capital markets. And, third, some of the heavy pipe-laying machinery to be used in the construction of the pipeline will provide the Soviet Union with access to Western technology in a fashion that will ultimately rebound to its strategic and military advantage.

On all three counts, the fears, while not groundless, are exaggerated. The most obvious danger is that Western European states which come to depend on Soviet gas supplies will then be uncomfortably in Soviet hands if, at some point in the future, the Soviet Union should decide arbitrarily to turn off the taps. The Soviet Union has, however, an enviable record for acting meticulously in the performance of commercial contracts, so that fears of strategic vulnerability presumably attach to the hypothesis that there might one day be a European war in which the Soviet Union was at loggerheads with its customers. Then, no doubt, the gas would be cut off but that would not bring any of the intending customers to its knees (although the war might). In no case is the amount of gas being purchased more than five per cent of the customer's use of energy, while the arrangements that prudent gas consumers will make to insure against a mechanical failure of the pipeline would no doubt also allow them to get by if hostilities should break out. Thus the chief strategic significance of the pipeline is that it would shorten the period of time during which Western Europe could fight a conventional war in Europe, and thus quicken the trigger of nuclear escalation. But this is entirely in conformity with the military strategy of Western Europe, not entirely welcome to the Soviet Union.

What the United States objections to the pipeline project overlook is the value to Western Europe of a supply of fuel which, although not dominant in the energy strategy of any potential customer, is at once relatively cheap and also secure. Not often, these days, can people hope to get their hands on a supply of energy over a period of three decades. And even the method by which the project is being financed, requiring that the customers should somehow help with the cost of construction, is not the simple gift it is made out to be. For who among the West German steelmakers and their workers, who are supplying pipes, or in the French companies from whom turbines have been ordered, will complain at the extra work that is falling in their laps? But is it not against the interest of the West that European states should make deals with the Soviet Union whose effects are to make the Soviet Union economically stronger? This is the unspoken question embedded in the American objection. The difficulty, of course, is that trade deals willingly arrived at do invariably benefit all partners. One of the continuing and deep-seated differences between the United States Administration and many governments of Western Europe turns on the European view that an economically healthy Soviet Union is a better neighbour than one perpetually on its knees.

The transfer of technology is in much the same case. Knowledge of how huge pipe-laying machinery works would no doubt be an advantage to the Soviet Union in the development of some of its remote fields of oil and gas, but that does not by itself imply that the world becomes a more dangerous place. The more serious of the United States' objections to what is proposed, however, is the decision that some of the high-capacity compressor turbines to be installed in the pipeline embody technology that could be of military value. This appears to be the reason why General Electric, the United States manufacturer, has asked its French licensee not to accept orders for the pipeline project. The decision is unlikely to make sense, for the same equipment can readily be had from elsewhere, for which reason it is hard to think it can be of crucial military value. But in any case, if President Reagan should give the pipeline project an airing in Versailles, he is likely to find M François Mitterrand, the President of France, taking a sharp but understandable line about the turbines — and the underlying principle.

Where now for UNEP?

On its tenth birthday the United Nations programme for the environment is stumbling.

The United Nations Environment Programme (UNEP), the principal offspring of the environmental jamboree at Stockholm in 1972 and a child of the 1960s, is in another bout of trouble. This week, the governing council of UNEP concluded its annual meeting in Nairobi, where it considered — among other things — the budget for the next two years. Predictably, there was a lot of heat and little light, but at least no delegation threatened to withdraw.

The budget is a constant worry, principally because of wavering by the United States, which has contributed $7 million a year since

0028-0836/82/220349-02$01 00

1973, more than a third of the total programme cost Last year's fear that the United States would contribute nothing at all did not materialize The State Department offer of $2 2 million was increased by Congress to $7 85 million — but UNEP is still waiting for the cheque This year, the Administration's offer is $3 million, but Congress may again vote more Whatever happens, and in spite of the surprise decision of the British government to increase its contribution by a quarter, UNEP's hope of doubling its budget ($17 million in 1981) now seems a dream So too must seem the first five years, when UNEP raised $60 million out of its target of $100 million within just a few months The peak budget ($21 million) was in 1977, since when, in spite of solid support from Scandinavia and the Netherlands, UNEP has gone downhill in strictly financial terms

In part, this is a negative consequence of the developing countries' growing attention to UNEP as a source of development aid One of UNEP's more successful exercises, the regional seas programme, began with Mediterranean pollution (which stems mostly from the French Rhône, Italian Po and Spanish Ebro rivers) but has now shifted to more southern seas This shift of attention is not just a UNEP quirk The environmental impact of poverty looms over that of affluence Growth is in fact faster in some countries of the developing world than in the "North", and where industrial development takes place, it is often without controls The air of Calcutta is now as bad as that of London in the smogs of the early 1950s, while London has clean air

Yet the Stockholm conference did not plan UNEP's $20 million a year to be an environmental action fund for the UN Rather, Stockholm saw UNEP as a coordinating agency within the UN system, drawing together and rationalizing the actions of the other UN agencies But the UN agencies — as always — proved jealous of their powers Moreover, the governing council of UNEP, which represents subscribing nations, ended up dealing almost exclusively with UNEP's small direct programme and budget, ignoring the wider role recommended by Stockholm

In the event, UNEP's programme comes to rest mainly on the Global Environment Monitoring System (GEMS), Infoterra (UNEP's information system), the International Register of Potentially Toxic Chemicals (IRPTC), and the Regional Seas Programme

The two last are generally thought to be the most successful They are based in Geneva, which gives them better relations with the other UN agencies (through the strong UN presence there), and they also happen to be headed by active and committed people Dutch biologist Jan Huismans for IRPTC, and Yugoslav marine biologist Stjepan Keckes for Regional Seas IRPTC is on the point of producing its first batch of complete data profiles of 330 chemicals of environmental significance — but it has been severely hampered by a halving of its budget (It relies now on Huismans, an assistant, and a secretary) Regional Seas is a little bigger (it has a professional staff of six), has fared better for funding with UNEP, and has scored a number of remarkable successes in the form of treaties controlling the pollution of coastal waters, notably the "Athens protocol" on land-based sources affecting the Mediterranean One reason for this success has been Keckes' decision to rely on local scientists — who often must be trained first — to assess pollution in an area, on the principle that governments are more willing to act on advice from their own scientists than on that of a fly-by-night international team The programme is now working on eight other regional seas around the world

Of the other programmes, GEMS — the global monitoring exercise recommended by Stockholm — has suffered from lack of resources A "colossal effort" is needed to strengthen this, says UNEP in a recent publication But an independent expert review of the 1972–82 decade, commissioned by UNEP, says that "even an imperfect series of environmental statistics would be an improvement on present ignorance" * And Infoterra, the information system, is barely used

UNEP's work on deserts — another great Stockholm concern — has also achieved little, not so much because of UNEP as because industrialized states have not wished to pay The UN Conference on Desertification held in Nairobi in 1977 produced good scientific papers and a plan of action, but the UN special account to combat desertification, set up after the conference, has remained almost empty By the end of 1981 total contributions stood at $5,000, all from a single contributor, Mexico

This is not the sum of UNEP efforts, however It can cite work on soils and pest control, its role in the World Conservation Strategy, which is seen as one of the successes of 1970s environmental management, reports on the impact of energy technologies, and a conference on renewables, its task force on pollution and human health (together with the World Health Organization), its promotion of a series of expert meetings on the preservation of tropical forests, international monitoring schemes for acid rain, and pollutants in food and fresh water The snag is that this part of UNEP's work is too much of a miscellany to make a mark

What next for UNEP? Apart from the cash problem, UNEP is concerned with defining its programme for the rest of the decade For the first time, the major UN agencies have got together to define a joint policy the "System-wide Medium-term Environment Programme" (SWMTEP) which covers the period 1984–89 But UNEP played little part in developing it, and it has been described as bland and highly generalized However, it distinguishes the roles that the different agencies should play, even if it is short on recommendations for specific action And it may help UNEP find its way through the UN thicket rather better than it did in the 1970s or it will, if this week's governing council gives UNEP the necessary backing

Universities in limbo

The British government's wish to save money on universities is forgivable; its way of doing so is not.

British universities now know that there will be no reprieve from the fate decreed for them by a thoughtless government The announcement last week by the University Grants Committee (*Nature* 27 May, p 258) that university budgets for 1982–83 will be essentially those forecast a year ago will have dispelled hopes that the government might not be serious The result is that contraction must continue for another two academic years There are two views of what the system will then be like The government seems ideologically committed to the belief that leaner means fitter Universities will have been forced to cut what is called "fat", and some will also have been forced to keep essential activities alive with help from local industry that will reduce demands on the public purse and be worthwhile in themselves What can be wrong with that, the government asks?

Plenty, is the simple answer The second and more gloomy view of the future starts from the observation that the contraction of the university system has been ordained within arbitrary constraints Last year's distribution of grants and, more important, the assignment of targets for student populations, has compelled some universities to shrink more rapidly than others And the restrictions on student numbers mean that no university can hope to work its way out of trouble by being more efficient — taking more students Yet demand for higher education in Britain continues to increase, with the result that students are shunted arbitrarily towards other institutions, the polytechnics for example All this arbitrariness arises, however, not from a belief that restricting access to higher education is in itself virtuous but because the government has not had the wit to devise a scheme for limiting the cost of maintenance grants for students, which are paid by local authorities and charged back to the central government

To seek to economize is within the government's gift, and may even be its duty To attempt to do so without solving a simple book-keeping problem, and while barriers persist between the two halves of higher education, amounts to maladministration

* See *The World Environment 1972–82*, edited by Martin Holdgate and others (Tycoly International, Dublin, 1982)

0028-0836/82/220350 01$01 00

Clouds may lift over Brookhaven

Congress moves to reprieve ISABELLE plan

Washington

ISABELLE, the new particle accelerator at Brookhaven National Laboratory on Long Island, may be saved for at least another year. Recently, serious problems arose with the original magnets, necessitating major changes in design, and causing delays and a rise in the project's construction cost from $275 million to $500 million. And because the troubled ISABELLE began taking a bigger than expected bite out of the US high-energy physics budget the question of whether to finish it at all has become a burning issue among US physicists.

It is also a burning issue in Congress, which is sensitive to pleas from the Long Island community to rescue the big machine, and, by extension, the Brookhaven Laboratory itself. The President's 1983 budget request included no funds for the construction of ISABELLE, although approximately $24 million was included for work on alternative magnets and other research and development projects. Moreover, a high level panel representing the national physics community had just said that ISABELLE should not be finished unless the overall budget was increased. So ISABELLE seemed doomed.

Or so it seemed until 7 May, when Congressman Tom Bevill (Democrat, Alabama), chairman of a key House appropriations subcommittee which has jurisdiction over the high-energy physics budget, toured the buildings at Brookhaven and declared himself "impressed". Better to see the creature, he said, than read about her in reports. Bevill assured Congressman William Carney (Democrat, New York) that he would try to get as much as $10 million in construction funds added to the budget when it comes before his subcommittee. Both Carney, who represents the Brookhaven area, and Thomas Downey (Democrat, New York) want to see ISABELLE finished almost as much as the scientists at Brookhaven do.

But ISABELLE has received only a qualified endorsement from other quarters. Congressman Don Fuqua, chairman of the House Science and Technology Committee which advises on how high-energy physics money should be distributed, has recommended either proceeding with ISABELLE or building a less costly machine at Brookhaven and using the ring tunnel already built.

Fuqua's recommendation carries some weight with Congress, as his committee knows as much about science as any on Capitol Hill. Although it may be construed as an endorsement for ISABELLE, in reality it steers carefully among the factions that have emerged on the question.

More guarded still was the report of a panel headed by George Trilling of the University of California at Berkeley, appointed to evaluate future plans for high-energy physics. Possibilities considered included proceeding with ISABELLE, cancelling ISABELLE to keep other machines going at other laboratories, and building something else. One of the options was a lower cost machine at Brookhaven, and in January the Trilling panel endorsed this recommendation when it reported to the High Energy Physics Advisory Panel of the Department of Energy. Thus, the Trilling group's view carries the weight of the community at large.

The Trilling report said that ISABELLE should be completed only if funds for high-energy physics in the future "averaged" better than the 1982 level of $395 million, for then ISABELLE's increase would not hurt other projects. However, since the President has requested $429 million for 1983, which is equal to the $395 million level in 1982 dollars, then, according to the advisory panel reasoning, ISABELLE should not be built.

All concerned are trying to avoid an outright showdown between the major US accelerator communities involved in the scramble for funding commitments for future big machines. The issue, they all say, is US leadership in high-energy physics in the 1990s.

The argument is partly about what kind of physics should be done, partly about how experiments are shaped by the machines themselves and their costs and timetables, and partly about keeping the eastern portion of the country, in which Brookhaven is the only remaining big accelerator facility, on a par with the facilities in the Mid-West and the West. Congressman Bevill, it seems, is proposing to resolve all this with a stroke of his pen when the item comes before his subcommittee in the next month or so.

Deborah Shapley

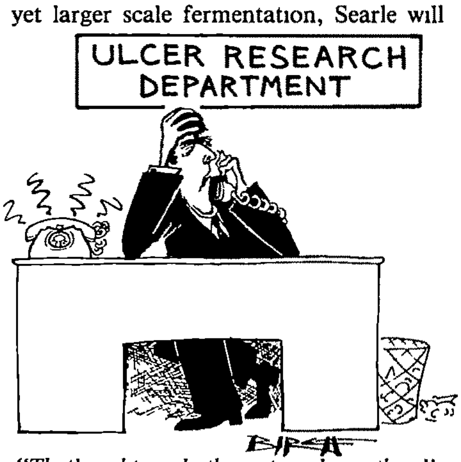
Ulcer cure sweetener for Searle

By the end of this year G. D. Searle and Co. Ltd expect to be testing their first recombinant DNA product in humans. It will be a peptide synthesized according to the instructions carried by a synthetic gene placed inside bacteria. Searle and Imperial Chemical Industries Ltd, initiators of the joint project, hope that the peptide, which is closely related to a gastrointestinal hormone, will prove to be an effective inhibitor of gastric acid secretion and hence of ulcers.

That is one indication that Searle's early attempts to capitalize on biotechnology have survived two setbacks. The first was the departure, in 1980, of two key members of staff to help start up Celltech. Searle's second setback was its failure to make a commercial success of producing beta-interferon from cultured fibroblasts.

Another indication of Searle's resilience in biotechnology came last week with the opening of a new £7 million biotechnology pilot plant at their High Wycombe site in the United Kingdom. The plant will come into operation next month, less than two years after construction started, despite a change in plan half way through. The change involved increasing the part of the plant designed for bacterial culture at the expense of that for cell culture. That shift in balance reflected both the failure of the beta-interferon project and, according to Brian Richards, vice-president of UK pre-clinical research and development, the discovery of several peptides of potential value that could best be produced by genetically engineered microorganisms.

The pilot plant still has facilities for the culture of cells but Searle anticipates more activity in the bacterial fermenters, the largest of which has a 450 litre capacity. The plant, however, is designed to accommodate a 4,000-litre fermenter, large enough to advance from pilot to manufacturing scale in some circumstances. For yet larger scale fermentation, Searle will



"That's right, we both create and cure them"

have either to build new plant or to arrange to use the vast fermentation facilities of the Japanese company Meiji Seika, using a recent agreement to exchange technology.

That, however, is for the future, when and if Searle has something worth fermenting. Bacterial products planned, apart from the anti-ulcer peptide, include interferons. Much effort is being directed towards making structural variants of interferons with greater or more specific activity than the natural compounds. If

0028-0836/82/220351-02$01.00

any are discovered the plan is to manufacture them with bacteria containing synthetic genes

Searle may also use biotechnology in the less glamorous pursuit of producing, more cheaply than at present, the two amino acids that are fused together to make aspartame, Searle's low calorie sweetener which was approved for tabletop use in the United States last year Both amino acids come from bacterial fermentation and so it is possible that the bacteria could be genetically manipulated to produce either amino acid more efficiently than at present **Peter Newmark**

Remote sensing

What war use?

Landsat, the US remote sensing satellite launched by the National Aeronatics and Space Administration (NASA) could have been used by the Argentine and British governments to obtain images of the Falkland Islands, but probably to little effect While Landsat can pick up data from all over the globe, it passes over the Falklands only every eighteen days and a further twenty-four hours are then required to process images Landsat has a resolution of 80 metres and could pick out — with luck and favourable weather conditions — an object as large as an aircraft carrier

Landsat-3 is a near-polar orbiting spacecraft launched in March 1978 It carries visible and infrared wavelength sensors which send back images used in agricultural surveys, mineral and oil exploration and pollution monitoring NASA collects the satellite data at Goddard Space Flight Center in Greenbelt, Maryland and it is then sold by the Interior Department through the Geological Survey's EROS (Earth Resources Observation Satellite) Data Center in Sioux Falls, South Dakota

Fresh images can be obtained from NASA by prior arrangement but the usual time-lag is anything from one to six months Ground stations within a 2,000 km range can pick up real time broadcasts by radio link The British defence department's National Remote Sensing Centre at Farnborough, Hampshire, is licensed under ESA's Earthnet programme to tune into transmissions over Europe Argentina also has a ground station at Mar Chiquita (600 km north-west of Buenos Aires), but to obtain fresh images of the Falklands, it would have to make a special arrangement with NASA so that the data would not be wiped off the recorders on board, which are limited in the amount of information they can store

The delay in obtaining fresh images and the coarseness of their resolution makes the strategic use of Landsat dubious Landsat images may, however, have provided information about prevailing terrain conditions **Jane Wynn**

Nairobi environmental meeting

More and less

A Japanese environmental delegation is this week trying to recover its pride, after its proposal for a "Brandt commission" on the environment took a battering at a series of environmental meetings in Nairobi

Japan had proposed that a group of independent and widely respected personalities get together to produce a weighty document on the global environment to the year 2000 Much to the delegation's chagrin, however, the Group of 77, representing developing nations, proved solidly opposed to the idea The group wanted direct development aid — for re-afforestation, for example — and not simply another expensive report for the bookshelves

This was only one of the little battles which have coloured Nairobi life in recent weeks, during the three meetings which celebrated the ten-year anniversary of the Stockholm environment conference The meetings held were one for non-governmental organizations concerned with the environment, a "meeting of special character" convened by the United Nations Environment Programme (UNEP) to review the decade, and the general council of UNEP itself

Environmentalists have left the meetings with mixed feelings A lot of hard political and scientific lessons had been learned over the decade, said one, but there was still a consummate lack of will to do anything practical

The current governmental attitude was represented by the level of the delegations sent by most countries to the "meeting of special character" UNEP invited all heads of state, but in the end there were just three — Presidents Mobutu (of Zaire), Nemeiry (Sudan) and Arap Moi (Kenya) The United States sent its head of the Environmental Protection Agency (EPA), Anne M Gorsuch, who made a fine speech recording US environmental support in the 1970s Ms Gorsuch has been generally regarded as President Reagan's axe-woman at EPA, so her speech came as a surprise However, she left her cutting comments for a press conference — the US contribution to UNEP would fall from $7 million in 1982 to $3 million next year, she said James Buckley of the State Department added that the US government saw UNEP as a "catalyst" rather than a prime actor

One bonus for UNEP, though, came from a surprising quarter Britain's junior environment minister Tom King promised an increase in the British contribution from £600,000 this year to £750,000 next Even allowing for inflation, this is a substantial increase, which the UK Department of the Environment must find within its cash-limited budget "We are conscious that resources are strained", said Mr King, and "now is the time for action"

Libya also promised its first ever contribution to UNEP — $1 million The Netherlands offered another 50 per cent, Japan, Finland, Malaysia, Uganda and Thailand also promised more But these increases would not cover the big cuts threatened by the United States

There was agreement on one thing, however a declaration, now to be known as the "Nairobi declaration", sixteen pages of fine prose "The world community of States solemnly reaffirms its commitment" says the declaration "to the Stockholm Declaration and Action Plan It also reaffirms its support for strengthening UNEP as the major catalytic instrument for global environment cooperation " The word "catalytic" is to be noted **Robert Walgate**

Information technology

Europe wakes up

Brussels

Europe's leading electronics and information technology companies are taking seriously the European Commission's grandiose plans to pool all their research efforts Dubbed "Esprit", the European Strategic Research Programme in Information Technology, it has evolved in a series of meetings that the Commission has held with leading European electronics firms during the past year

The first fruits of these discussions have been leaked from a communication sent to the Council of Ministers in preparation for the EEC's Science Council to be held on 30 June Yet despite industry approval, it is in the political arena that the real fight will take place to get the Commission's ambitious ideas implemented Discussions with UK government bodies such as the Department of Industry and the Science and Engineering Research Council have left officials in Brussels with the impression that although there is token acceptance of Esprit's principles, the national officials doubt that they can ever be realized

Europe is losing out to Japan and the United States in the race to develop information technology, argues the Commission, not for want of spending vast sums on research Siemens alone devotes around $800 million a year to research Yet Siemens, ICL and the others are in financial difficulties and failing to reap the rewards from their research investments

Industrialists feel strongly that to get the best out of Europe's research expenditure a new body should be set up to coordinate activities Far from wishing to build empires, Commission officials dismiss the idea that they themselves should tackle this A more professional body such as the United Kingdom's National Research and Development Corporation with experience in turning research into marketable products would be given the job It would be co-financed by the companies involved

0028-0836/82/220352-02$01 00

and the member states, and topped up from the EEC's budget.

The management of Esprit would be decentralized and controlled by industry. The aim would be to copy the Japanese by setting clear strategic objectives and concentrating research and exploitation on a few large companies. Areas already singled out for concentrated exploitation are: advanced microelectronics; advanced information processing systems usable by the man in the street; systems for office automation and computer integrated flexible manufacturing; and fourth and fifth generation computers. The Commission is also asking for some $11 million to spend on preliminary pilot projects to sort out the best ways of implementing cooperation.

Radical though the scheme is, the Commission is pressing for an agreement by December 1982 and a start in January 1984. Without such a European effort being launched quickly, officials in Brussels are saying, European companies will increasingly be forced to take minor roles in cooperative deals with Japanese or American competitors — in the United Kingdom ICL is already moving in this direction.

While the Commission may be accused of being over-ambitious, it is clear that Esprit is being taken very seriously by Plessey, GEC, ICL, Nixdorf, Siemens, AEG, Honeywell-Bull, CIT Alcatel, Thomson-CSF and Philips. **Jasper Becker**

Space research

UK goes national

For the first time in ten years, Britain is trying to devise a coherent space policy. The impetus has come from the recognition, later than in other Western countries, that there is money to be made from selling space technologies and services. The British government's apparently sudden decisions, earlier this year, to allow and even encourage direct broadcasting by satellite and cable television point to its fear of losing out on new technologies now coming of age elsewhere in Europe and which are already well established in the United States. The plans now being discussed for a national remote sensing programme, based on the European Space Agency's recently approved Earth Resources Satellite 1 (ERS1) are thus meant to prepare British industry for the time when money can be made from remote sensing data and hardware.

But if the spirit has changed, where will the money come from? Public money will not flow freely, so what about private investment? A select group of financiers and representatives from the space community met last week in the Surrey country house belonging to Logica, the software company, to address that question. Disappointingly the sound of money changing hands was muffled.

What little space policy Britain has had has been directed through its membership of the European Space Agency at the space segment, the hardware in orbit. Now it seems to be agreed that more attention must be paid to the use of data sent back from space — and to persuading people that they must pay for it.

Financiers, still wary of the risks involved in satellites, see more scope for helping to market space on the ground.

Based on the experiences of direct broadcasting and cable in the United States, they are convinced that there is a market in Britain, while industry seems keen to produce antennae and dishes for receiving direct satellite broadcasts. What the financiers want to know is how big the market will be.

Remote sensing, for which the market is diffuse and ill-defined, faces different problems. A programme of scientific research using ERS1 data has been worked out by the Natural Environment Research Council and the Science and Engineering Research Council, but nobody yet knows who will pay for remote sensing data. According to Sir Hermann Bondi, chairman of the Natural Environment Research Council, Britain is fortunate that ERS1 will be an oceanographic monitoring satellite. The need to analyse ocean data quickly because its usefulness is often only immediate will, he says, pave the way for the later use of land data.

Land remote sensing data may be more saleable, perhaps to mining companies and agriculture, but only in the long run. The experience of the National Aeronautics and Space Administration in running Landsat is, however, unhelpful; the early promises made for the data have not been fulfilled. The meeting last week seemed, however, confident that data from Landsat D, to be launched later this year, and from Spot, the French national remote sensing satellite, will be more attractive.

One problem that perplexed the meeting is that the applied science of remote sensing falls administratively between two stools, the Science and Engineering Research Council (science) and the Department of Industry (application). This is why there has been a spate of interest in the question whether Britain should have a space agency of its own. Given departmental jealousy, the outcome will probably be some kind of coordinating agency, more modest, for example, than the French.

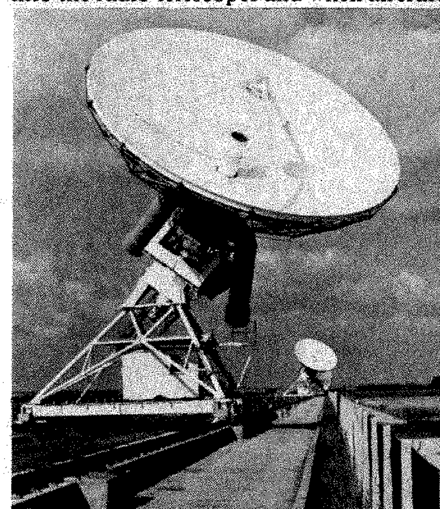
Judy Redfearn

British radioastronomy

Fear of flying

The British propensity to jetset may cause acute problems for Cambridge radioastronomers. In the next few months Graham Eyre, QC, the inspector at the public enquiry into the proposal to site the third London airport at Stansted in Essex, will hear evidence from the University of Cambridge that an expanded airport at Stansted would seriously impede work at Mullard Radio Astronomy Observatory.

The Mullard Observatory, which is situated some 35 miles from Stansted, has for many years had an agreement with the Home Office whereby aircraft interference is controlled. Interference occurs when ground transmissions are reflected back into the radio telescopes and when aircraft equipment gives out radio waves on the telescope's frequency.



In-line for interference — the 5-km radio telescope.

If Stansted were to be expanded to take 15 million passengers a year, the option favoured by the British government, the increase in air traffic would mean that aircraft approaching the airport would start their landing manoeuvres within 4 to 5 km of the observatory. Work using low frequency wavebands — observations of the solar and interplanetary medium and of supernova remnants, for example — could be completely disrupted.

In support of its evidence to the enquiry, the observatory will also present a technical document assessing the amount of interference and the consequences for research, which are not easily quantified. The problem is that the frequency bands allocated to radioastronomers are also shared by other radio users, at present not allowed within 50 km of the observatory. With aircraft, the range would need to be expanded to 200 km, which is not a practical option for a busy airport.

The observatory has occupied its present site since 1956 and the area, which was selected for its lack of radio interference, has been kept "clean" ever since. It would be expensive and difficult to find another suitable site. **Jane Wynn**

0028-0836/82/220353-01$01.00

Kibbutz biotechnology

Farming today

Rehovot

The traditional view of the Israeli kibbutz — hard work in the fields under a blazing sun — may have to be changed. For the future of many kibbutzim may lie with small science-based industry using the techniques that are already spawning new biotechnology industries in the United States and Europe.

The kibbutz at Beit Haemek is typical of the trend from farming village to light manufacturing to science-based industry. In the mid-1970s, Beit Haemek was looking for a way of supplementing its income from the cultivation of avocados and the raising of cows. And when it failed in an attempt to balance its budget by building and selling pipe organs, the kibbutz decided to establish a company that would use advanced tissue culture techniques to propagate ornamental, specific pathogen-free plants. This move made sense as the kibbutz had several members with degrees in agriculture and microbiology, and others with scientific training and laboratory experience.

Further expansion of the enterprise, now called Biological Industries, was facilitated by the arrival of a new American member in what had been a predominantly British settlement. He is Dr Robert Levin, previously a research associate at Stanford University Medical School.

The kibbutz can do only a limited amount of research on its own, and so Levin works closely with Professor Jonathan Gressel and his group at the Weizmann Institute of Science. One result of this collaboration has been the exploitation at Beit Haemek of a new Weizmann Institute system to use plant tissue culture methods for rapid and inexpensive prescreening of potential herbicides. And in cooperation with scientists at other Israeli research centres, Levin and his colleagues are now utilizing such methods to develop plant varieties resistant to herbicides, diseases and other environmental hazards.

Each successful new project has brought an expansion of Biological Industries and so now Beit Haemek has caused some surprise in Israel by placing advertisements in the newspapers inviting biologists with PhDs to join the kibbutz and applicants are now being screened. The introduction of such specialists would be quite a change in traditional kibbutz practice. On the one hand, it will be necesssary for veteran members of the kibbutz to accept the fact that the newcomers will not be available, as kibbutz tradition decrees, to do any job necessary; they will expect to work in their professions. On the other hand, the new members themselves will have to come to terms with a society where there are no material rewards for their degrees or for their specific contributions.

Yet the egalitarian economic set-up of a kibbutz does seem to be consistent with making a success of a sophisticated science-based industry. Beit Haemek's enterprise expects a $1 million turnover this year, with 65 per cent of its tissue-culture-derived products being exported, mainly to the Netherlands and West Germany.

The kibbutzniks see themselves as a production arm of Israeli academics, in the words of Beit Haemek's general manager Mike Landis, "as a conduit between Israeli research centres and the world market".

Robert Levin of Beit Haemek sees a great deal of similarity between the science-based industries on kibbutzim and the new biotechnology companies in the United States, "which are", he says, "really cooperatives created by a few talented scientists. The only difference is that kibbutz scientists in Israel share modest living quarters while capitalistic scientists in the States share enormous profits".

Nechemia Meyers

Nature in Poland

Many readers will have been moved to ask how they can help by last week's report of the difficulties encountered by Polish scientists in acquiring scientific journals from abroad (see *Nature* 27 May, p.259). The chronic shortage of hard currency, which has apparently worsened since the introduction of martial law last December, has made the purchase of journals virtually impossible.

In the belief that many scientists outside Poland may wish to help colleagues thus deprived, *Nature* is introducing a scheme whereby scientists outside Poland may purchase a year's subscription to *Nature* in Poland at roughly half the usual price.

Those wishing to buy a subscription for a Polish colleague should write to *Nature* in either London or New York (the addresses are given opposite page 349), enclosing a cheque made payable to *Nature* for £50, or the equivalent in any currency, together with their full name and address and those of the intended recipient. Telephone orders may be charged to international credit cards. *Nature* will meet the extra cost of despatching the journal to Poland.

Immediately on receipt of an order, the intended recipient will receive a letter from the Editor describing the arrangement that has been made and identifying the donor. As soon as the order has been processed, copies of *Nature* should arrive in Poland within a week of publication. This offer will remain open at least until 1 August.

Grenoble neutron beam reactor

Working again

Europe's high flux neutron beam reactor, the Institut Laue-Langevin (ILL) at Grenoble, is likely to re-start in June with an unrepaired hole in its primary cooling circuit.

Admittedly, the hole is inside the heavy water primary cooling circuit, and is not considered dangerous, but careful experiments and calculations are being made to make sure that the reactor will work safely despite the fault.

The broken part is a section of perforated aluminium alloy, designed to break up the flow of the cooling water before it turns a corner in the piping. A 250 cm^2 section broke away, and parts of this interfered with the operation of a cooling pump. As soon as the fault was discovered, the reactor was shut down, and already nearly a month's experiments have been lost — a lot of work at ILL, which services dozens of visiting groups working simultaneously. The groups use the neutrons, often for only a few days, but the data so taken provide much more work on data analysis back at the group's home institutions.

Dr Brian Fender, ILL's British deputy director, said last week that he expected to be able to save the 1982 experimental programme by putting off a scheduled development period, due in November, until next February. Ultimately, some two months' experimental time at ILL will have been lost; but it will be somewhat less painful if the loss is taken next year, since 1983 experiments are still only in the planning stage.

The aluminium part broke because of ten years' cycling in turbulent water flow, ILL engineers believe. Well away from the reactor's fuel rod, it did not suffer significant radiation damage. Almost all the pieces have now been recovered. Reassembly of the jigsaw has shown a piece a centimetre across still to be missing, presumed lodged in the cooling circuits, but it is not expected to do more damage when the reactor is restarted.

Flow through the cooling circuit has been tested, using a dummy fuel element, and despite the hole in the aluminium section, no increased turbulence or vibration has been observed. The broken part has also been inspected by TV camera, from a distance of one or two metres, and seems firm enough. It is conceivable that the French licensing authorities would not allow ILL to restart without replacing the part, but, says Fender, "we work very closely together", and he does not expect any trouble from that quarter.

Why could the broken part not simply be replaced? Because it is awkward to reach, requiring the dismantling of most of the reactor, an exercise which could cost many more months than the present delay.

Robert Walgate

0028-0836/82/220354-02$01.00

CORRESPONDENCE

A letter to Soviet scientists

The following document reached *Nature* on 21 May 1981, the 61st birthday of Academician Andrei Sakharov.

I have learned from foreign broadcasts that the "Scientists for Sakharov, Orlov and Shcharanskii" (SOS) Committee in the United States has issued an appeal to Soviet scientists to come to the defence of the family of Sergei Kovalev — that is, Sergei himself, his daughter-in-law Tat'yana Osipova, and his son Ivan. I feel sure that many of you also heard this appeal. I beg you to treat it with all seriousness and to undertake those actions to which each of you feels prompted by his feeling of responsibility as a citizen and by his conscience, both in this particular case, and in others where your intervention would be of special importance.

For many years now, I have observed an almost total lack of action on the part of my Soviet scientific colleagues in matters relating to the defence of human rights. It is shameful that foreign scientists (and deepest thanks to them!) show a greater concern for our affairs than we do ourselves. In this letter, I am speaking first and foremost to those scientists occupying a fairly high and independent position — more specifically to Members and Corresponding Members of the Academies. I know that some of you have negative or sceptical feelings regarding my statements on general questions. But this is not a general question (on these, I have always expressed my views in the form of a discussion, and have never thrust my opinions on to anyone). This is a matter of the fate of real people, including the fate of your colleague in science. Every one of you who has shown even the least interest will form his or her own independent opinions on these cases, of the total degree of injustice and the harshness of repression.

You cannot consider that these cases do not touch you; the historical experience of our country, professional solidarity, simple human sympathy for the fate of others, and often personal contact with the victims of repression (many of you were personally acquainted with Orlov, Kovalev and others), make such an attitude impossible. You must not invoke the interests of your job, the need to keep your professional post in the interests of science. In fact, most of you are in a fairly secure position. And the interests of science must include the defence of members of the scientific community from injustice; they include the responsibility of the citizen. This is not the Stalin era; practically speaking none of you *now* are under any kind of threat. But can you rule out completely a return to a new era of mass repression? For if this should come to pass, your lack of action *now* will be one of the causes. And conversely, could not the activity as citizens and the independence of even a few of the country's top scientists have a profound and beneficial effect on the whole situation?

And which of you can lay his hand on his heart and proclaim that this is not a vital necessity? It is precisely this which more than anything would promote international scientific relations, and, on a far far wider scale, trust and peace throughout the world. Such is the measure of the individual personal responsibility of every one of you.

In conclusion, I should like to cite several examples of scientists and scholars who are victims of repression — prisoners of conscience. For each of these cases there is information which is accessible to you.

(1) The case of Kovalev and his family, with which I began this letter. Kovalev has served a term of seven years imprisonment under the harshest conditions, and he is now in exile, under shocking conditions in the Magadan region. Tanya Osipova and Ivan Kovalev are in prison, and it is necessary to obtain for them, at the very least, the opportunity of meeting. After the arrest of his wife, and before his own arrest, Ivan had only one 30-minute meeting with her in 15 months.

(2) The case of Yurii Orlov. His fate, and the cruel torments he is suffering are widely known.

(3) The case of Anatolii Shcharanskii — also very well known.

(4) The case of the mathematician and cyberneticist, Dr Aleksandr Bolonkin. After imprisonment and exile which robbed him of nine years of his life, beatings-up, confinement in the punishment cells and provocations, he is back in prison again on a trumped-up criminal charge.

(5) The case of the philologist and poet, Vasyl Stus'. After many years in prison and exile, he was condemned to a further 15 years labour-camp and exile, simply for supporting the Ukrainian Helsinki Watch Group.

(6) The case of Anatolyi Marchenko. He is not a scholar; he is a worker and writer, the author of some remarkable books on present-day labour camps and prisons. But he has been sentenced again — his fifth sentence — to the monstrous term of 15 years imprisonment and five years exile; the central point of the charge being his open and noble-hearted letter on the "Sakharov case" to academician P. L. Kapitsa, who is one of those to whom I address this letter.

(7) The cases of Aleksandr Lavut and Tat'yana Velikanova (mathematicians), Mart Niklus (ornithologist), Leonard Ternovskii (radiologist), Viktor Nekipelov (a pharmacist and talented poet), Pyatkus (philologist), Meilanov (mathematician), Luk'yanenko (jurist), Airikyan, Altunyan and many others. None of them has had recourse to violence nor would do so. All of them have suffered cruelly for loyalty to their noble convictions, the principles of free expression and justice. Help them — this is our common duty.

Andrei Sakharov, Gor'kii, 30 March 1982.

POSTSCRIPT — **In defence of Ivan Kovalev.**

Ivan Kovalev has been sentenced. This is yet another blow to the champions of human rights in the Soviet Union, to free expression, and to the Helsinki Watch Group; a blow against the principles expressed in the Helsinki Final Act. And this is a new act of inconceivable cruelty towards the famous Kovalev family, the third of its victims. In December 1974, Sergei Kovalev, Ivan's father, was arrested. In May, 1980, Tat'yana Osipova, Ivan's wife was arrested. Both of them, like Ivan Kovalev himself, have received cruel sentences for their nonviolent defence of human rights, for their fidelity to noble convictions. In a letter which he wrote shortly before his arrest, Ivan Kovalev wrote about himself and his situation. And in his article "My Tanya", he told us about his wife, her fate and her struggle. There is little one can add to those documents — and having known Ivan for many years I am convinced of the justice of his case.

I call upon everyone who believes in justice, who can feel for the tragedy of this young family, to speak out in defence of Ivan Kovalev, his wife and, of course, his father. The possibility of a meeting between Tat'yana and Ivan must be obtained — this is their inalienable human right. And it is very important too to obtain their freedom, and the freedom of all prisoners of conscience.

Andrei Sakharov, Gor'kii, 2 April 1982.

ESA clarification

SIR — May I refer to your article "Antarctic research hit by crisis" published in *Nature* of 15 April (p.593) and in particular to the part of the article:

> "Although Argentina's own research efforts have been extremely limited, it is at present collaborating with the French in glaciological work, and plans exist to site a ground station for the first European Space Agency remote sensing satellite to be put into polar orbit, at the Argentinian base of Marambio".

There is no Argentinian connection with the European Space Agency/ERS-1 programme. There will certainly be some ERS-1 overseas ground-stations outside Europe, but no negotiations have been made with Argentina or indeed with any other South American state. If such plans were ever envisaged, they would require approval of ESA Council.

You will appreciate ESA's desire to have this point clarified.

W. BRADO
(Head of ESA Director General's Cabinet)
ESA, Paris, France

Editor in false garb

SIR — If both of your bizarre and intemporate recent leading articles ("Doves in false garb", 18 February, p.542; and "Professional propaganda", 1 April, p.380) had been published on 1 April, one might more easily have understood them. They are so full of rancid sarcasm, repetition to the point of perseveration, wild claims devoid of any supportive evidence, multiple errors in basic logic, frank contradictions and odd grammar, that they imply a writer almost incoherent with anger. They are also probably the most disgraceful abuses of editorial privilege I have ever seen in print.

You have used, however ineffectually, the anonymity of the editorial and the prestige of a hitherto distinguished scientific journal, to lend spurious respectability and authority to your personal and highly idiosyncratic views. You have jeered at a very large number of sincere, ethical, distinguished and honourable professionals throughout the world, impugning their honesty, without raising a

single justified or logical argument. You misrepresent the views and claims of those you disagree with, so as to more easily oppose them.

You criticize the professions (with biased inaccuracy) for ignoring public issues, and belittle them when they speak out on them. You graciously grant us the responsibility to contribute our special knowledge to the solution of such problems, but criticize us for failing to exceed the limits of such special knowledge. You insist that it's laudable for us to express ourselves singly, but dreadful to do so collectively. You sneer at the "catchy acronymic titles" of the groups concerned, though few of them are catchy, and none are acronyms. You complain that the groups "claim professional support" in a deceitful way. While the degree of professional support they have is significant and growing, their titles very specifically and accurately state their membership as limited to those members of each respective profession who are opposed to nuclear arms or war: how much more accurately could they describe themselves?

Take your own advice, Mr Editor. Put aside your pique about the professions you don't belong to; put aside your anonymity and the authority that belongs to *Nature* and not to you — take your courage in your hands as you exhort us to do (and as we have done) and join us on the hustings. If you can express your views coherently and support them with data, we will listen with interest and respect.

Amongst your hectic hubris, you graciously concede that "it would . . . be shocking if physicians were silent about the probable consequences of smoking". May we not also warn about the dangers of being smoked?

Michael A. Simpson
Health Sciences Center, Temple University, Philadelphia, Pennsylvania, USA

Ab what?

Sir — In the preamble to the first editorial in the edition of *Nature* dated 22 April (p.693), the word "abreact" is used. So far as I can discover the word is a term from psychoanalysis referring to the liberation of the individual from neurosis by the expression of repressed emotion. Is this what the leader article intends?

I feel it would be a pity if *Nature* were to start falling into the trap of employing jargon terms made meaningless by fashionable use.

It might cause some adverse reaction among the readers.

Robin H. C. Strang
Department of Biochemistry, University of Glasgow, Glasgow, UK

Data protection

Sir — I read with great interest the article "Big Brother's law" (*Nature* 1 April, p.694) about personal data protection. I would like to add a few comments.

In the summer of 1981, the European Commission addressed a formal recommendation to European Community member states to sign and ratify the convention of the Council of Europe by the end of 1982. Up to now, only five member states have signed it.

In this recommendation, the Commission stated that it would draft a Community directive if the member states failed to sign the convention.

The socialist group in the European parliament has long pressed for a European Community directive. Indeed, in March 1982 the European parliament adopted a report of its legal committee (the report was given by Herr Sieglerschmidt, SPD) asking among other things for a Community directive for personal data control.

We believe, however, that the most important point is the control of the transnational data transfer of private companies, although you are right in the statement that the transfer of information between governmental departments is very important.

Finally, I would like to say how important it is to lobby the British members of the European parliament. The international companies are very familiar with effective lobbying of MEPs. The electorate should do the same, but in the interest of maximum protection of data on individuals.

Fritz Gautier, MEP
Braunschweig, FRG

Patent sense

Sir — The report in *Nature* of 29 April (p.792) of the 21 April Colloquium at the Association of the Bar of the City of New York is inaccurate in stating that a European visitor may, under European laws, derive an invention from a seminar or talk in America and appropriately apply for a patent in Europe.

European and other patent laws do not permit patents to those who derive. Typically, they require that the subject of a patent be the original work of the applicant or its predecessor in interest.

The issue of whether or not an invention is original work is different from the issue of whether or not the inventor's (original) work preceded the work of another.

David W. Plant
Fish & Neave, New York, USA

Ageing controversy

Sir — I am writing to vehemently protest about the recent *News and Views* article from your ageing correspondent in the 1 April issue of *Nature* (**296**, 392-393;1982). The work of Dr Obispo and his colleagues on "longevin", the lifespan-extending protein from carp gut is described in excruciating detail, but no mention at all is made of Dr A. Huxley's pioneering work.

It was in the obscurity of his Oxford college rooms that the re-discovery of the rejuvenating effect of raw fish viscera was made (*Phil. Trans. R. Soc.* B **240**, 153-215, 1969). Furthermore, he has gone a long way in the characterization of the same protein. I am truly surprised that this work of a British scientist (who in addition comes from a very distinguished family indeed) should be overlooked in the pages of *Nature*. I think the reason for this is not only that the Obispo and Maunciple group has essentially unlimited funding from the Stoyte Foundation for Aging Research, but also that nowadays anything having to do with gene cloning receives immediate attention, while people like your correspondent tend to overlook standard protein chemistry work. It is sad that Dr Huxley's work has been ignored for so many years, but it is even more sad that the California group should be allowed to patent the longevin B-chain clone. The extension of human lifespan is of utmost ethical importance and should not be treated lightly; least of all in a patent office.

An Ageing Professor
Basle, Switzerland

Sir — I read the *News and Views* article on longevin (*Nature* **296**, 392-393; 1982) with keen interest but think, if I may say so, that this may prove to be the swan song of geriatrical biochemistry.

Jennifer Reed
Institute for Experimental Pathology, German Cancer Research Center, Heidelberg, FRG

Sir — Huxley in his seminal work *After Many a Summer* (Chatto and Windus, 1937) was the first to draw wider attention to the research of Obispo. It is regrettable that your correspondent's otherwise excellent report on anti-senescence factors (*Nature* **296**, 392-393; 1982) did not refer to Obispo's pioneering studies, even though these were conducted on whole carp *flambé* (*J. Carp Sci.* **1**, 69; 1936). Hauberk's results on raw carp viscera were not rediscovered until the regular visits of a member of the present shadow cabinet to Tashkent led to the realization that the famous mad ape in the local zoo was none other than the now fully evolved Lenin. Obispo himself, his procreative and creative powers equally unimpaired, will reach the age of 100 in 1984, by which time, we can be confident, gerontology will have experienced a timely death.

Richard Glendale
London W6, UK

Nuclear options

Sir — David Fishlock should keep his journalistic innuendos (*Nature* 22 April, p.700) for *Financial Times*. He implies that my letter in *Nature* of 18 March (p.192) on "Electricity costs" is equivalent to flat-Earth theory, ostensibly because it cites a paper of mine[1] as "in the press". He calls this "unpublished", but in fact Energy Policy No. 2, containing the paper, was published on 5 May, and the *Financial Times* was sent a copy of the page proofs before publication. In any case the scientific editor of the *Financial Times* was surely not unaware of the report of Sir Kelvin Spencer's committee[2], which was given wide publicity after a press conference in the House of Commons on 2 February, and which quoted or paraphrased the main results of the "unpublished" paper.

I can only assume that any conclusion which does not wholeheartedly support the nuclear case is necessarily flat-Earth theory to David Fishlock, and need not be looked at any further.

J. W. Jeffery
Department of Crystallography, Birkbeck College, London WC1, UK

1. *The Real Cost of Nuclear Electricity in the U.K.*, *Energy Policy*, Vol. 10, No. 2, 76-100 (1982).
2. *Nuclear Energy: The Real Costs; A Special Report by the Committee for the Study of the Economics of Nuclear Electricity.* (Worthyvale Manor, Camelford, Cornwall).

Gravitational radiation and the binary pulsar

from Virginia Trimble

GRAVITATIONAL radiation not only exists, but it is apparently also doing exactly what general relativity says it ought to, at least in the binary pulsar PSR1913+16. This is the conclusion drawn by J.H. Taylor and J.M. Weisberg writing recently in the *Astrophysical Journal* (**253**, 908; 1982). It comes after more than six years of observing gradual changes in the eight-hour orbit of the pulsar and its invisible companion, probably also a neutron star.

The pulsar period itself (0.059 s) acts as the known emitted frequency from whose Doppler shifts the orbit is determined as part of a total solution including up to 20 variables. These represent intrinsic properties of the pulsar, classical orbital elements, and relativistic corrections for the advance of the perihelion (about 4° per yr compared with 43″ per century for Mercury) and variable parts of the gravitational redshift and the transverse Doppler effect. The relativistic parameters together define a small set of allowed values for the masses of the two stars. These, plus the orbit period, can be plugged into a standard formula for gravitational radiation from orbiting point masses to yield a unique prediction of the rate at which the orbit should be decaying, $\dot{P} = -2.4\times10^{-12}$. The measured value, from the multi-variable fit, is $-2.30\pm0.22\times10^{-12}$, in exact agreement to within the experimental errors.

The high precision and self-consistency of these results mean that they have interesting implications for three problems of current interest in astrophysics.

First, if no important effects have been left out, the binary pulsar orbit provides as accurate a mass for a neutron star as has ever been determined, and the first one for an honest-to-goodness radi pulsar. The other measured masses all come from X-ray binaries, for which the energy source is not the rotation of the neutron star (as in 'true' pulsars) but accretion of gas from a companion, an inherently messy process. The answer is $1.4\pm0.1\ M_\odot$ for each component. Existing data[1] are consistent with all known neutron stars having about this mass. It is just above the maximum stable mass (Chandrasekhar limit) for a star or stellar core supported by degenerate electron pressure and may reflect a fairly universal truth — you get a neutron star when an inert core builds up to the Chandrasekhar limit and so collapses.

Second, we are encouraged to ask more seriously than before how gravitational radiation will affect the evolution of other short-period binary systems, whose orbits are less clean than that of PSR1913+16, but which are astronomically at least as interesting. The classic case is that of the cataclysmic variables (CVs)[2-6] — novae and related systems consisting of a normal star transferring material to a white dwarf companion. Many kinds of binaries show such mass transfer. In some, the flow is driven from a more massive to a less massive star by newtonian effects; in others, a star spills over because it is trying to expand and become a red giant[7]. But in some low-mass CVs, neither of these explanations works. The shedding star is the less massive and has not yet begun to expand. What drives the transfer? Gravitational radiation must gradually bring the two stars closer together. The surprise is that it does so at precisely the rate needed to produce the otherwise inexplicable mass transfer in these short-period, low-mass CVs[8-10]. In addition, beyond a certain point, gravitational radiation-driven mass transfer begins to move the stars apart again, and so may also explain some puzzling gaps in the distribution of orbit periods of CVs.

Third, the binary pulsar provides a test of the standard formula, first derived by Einstein[11], for the gravitational radiation expected from a binary system. Over the years, many theorists have tackled the problem within the framework of general relativity. Some have confirmed the so-called quadrupole formula[12-14], whilst others have found different results, or, more often, concluded that a proper derivation has not yet been carried out and is currently beyond them, too[14-16]. The agreement to within 10 per cent between prediction and observation for the binary pulsar strongly suggests that the quadrupole formula is very nearly right, at least for relatively weak-field cases (again, unless some important effect has been left out of the multi-variable analysis). This should help theoretical relativity both with this particular problem (it is always easier to do a calculation if you know the answer in advance) and with others by suggesting appropriate approximation methods. General relativity is not the only theory of gravitation currently under investigation. But most of the others (except some cases of the Brans-Dicke theory) predict orbit changes due to gravitational radiation that disagree with the binary pulsar result[17]. We cannot, however, quite rule them out yet, if only because the calculations leading to that conclusion have been done with methods analogous to those that lead to the quadrupole formula for general relativity, which may, just possibly, not be appropriate.

What can we expect from PSR1913+16 and gravitational radiation in the future? Observations are continuing. And the nature of the problem is such that precision increases roughly with the square of the observing time. Thus another five years could narrow the error bars on the relativity parameters from about 10 per cent to about 1 per cent, providing neutron star masses with that precision and testing agreement (or disagreement) to 1 per cent between the quadrupole formula prediction and the observed value for the orbit decay due to gravitational radiation. Beyond this we probably cannot go. Fluctuations in the newtonian gravitational potential of the Galaxy (due to stars, clusters, gas clouds, spiral arms and all the rest) cause random, undecipherable accelerations of a test particle and so will put noise into the pulsar timing measurements at a level near 2×10^{-14} on time scales comparable with the effects being sought.

Finally, once a phenomenon has shown up indirectly in an astronomical context, it is natural to ask whether it can be seen directly in the laboratory. Detectors designed to search for gravitational radiation pulses have been in fairly continuous operation one place or another for

Virginia Trimble is in the Department of Physics, University of California, Irvine, California 92717, and the Astronomy Program, University of Maryland, College Park, Maryland 20742.

0028-0836/82/220357-02$01.00

about 15 years. Ferrari *et al.*[18] have reported a recent data set in which widely separated detectors were simultaneously excited by some sort of background and reference earlier results. Helium-cooled detectors of higher sensitivity are being developed or are in operation at Rome, Stanford, Maryland, Tokyo and other places. The expected gravitational radiation flux at the Earth from the binary pulsar is, however, only about 10^{-8} ergs $cm^{-2} s^{-1}$, less than 10 per cent that expected from the Crab Nebula pulsar, and some 10 orders below the sensitivity of foreseeable ground-based equipment at the relevant frequencies. Indirect methods will, therefore, clearly continue to be important for a long while, at least for PSR1913 + 16!

1. Bahcall, J.N. *A. Rev. Astr. Astrophys.* **16**, 241 (1978).
2. Kraft, R.P., Mathews, J. & Greenstein, J.L. *Astrophys. J.* **136**, 312 (1962).
3. Braginski, V.B. *Usp. Fiz. Nauk* **86**, 433 (1965).
4. Paczyński, B. *Acta astr.* **17**, 207 (1967).
5. Faulkner, J. *Astrophys. J. Lett.* **170**, L99.
6. Pringle, J. & Webbink, R.F. *Mon. Not. R. astr. Soc.* **172**, 493 (1975).
7. Paczyński, B. *A. Rev. Astr. Astrophys.* **9**, 183 (1971).
8. Paczyński, B. & Sienkiewicz, R. *Astrophys. J.* **248**, L27 (1981).
9. Paczyński, B. & Sienkiewicz, R. *Preprint* (Orange Aid Series, 1982).
10. Rappaport, S., Joss, P.C. & Webbink, R.F. *Astrophys. J.* **254**, 616 (1982).
11. Einstein, A. *Sber. preuss. Akad. Wiss. Phys.-Math* **K1**, 154 (1918).
12. Landau, L. & Lifschitz, E. *Classical Theory of Fields* (Addison-Wesley, Massachusetts, 1951).
13. Walker, M. & Will, C.M. *Astrophys. J. Lett.* **242**, L129 (1980) and refs therein.
14. Cohen, J.M. *Ann. N.Y. Acad. Sci.* **375**, 459 (1982) and refs therein.
15. Ehlers, J., Rosenblum, A., Goldberg, J.N. & Havas, P. *Astrophys. J. Lett.* **208**, L77 (1976).
16. Cooperstock, F.I. & Hobill, D.W. *Astrophys. J. Lett.* **235**, L55 (1980).
17. Weisberg, J.M. & Taylor, J.H. *Gen. Rel. Grav.* **13**, 1 (1981).
18. Ferrari, V., Pizzella, G., Lee, M. & Weber, J. *Phys. Rev.* **D25** (May 15, in the press).

Antibodies and cancer therapy

from P.C.L. Beverley

A RECENT report describes the dramatically successful use of a monoclonal antibody in the treatment of a human B-cell tumour[1]. The patient has now been clinically free from disease for eleven months without further treatment. This striking success was largely due to the exquisite specificity of the antibody for the tumour cells, but is something of a special case. The monoclonal antibody was anti-idiotypic: that is, it was raised against specific determinants on the antibody expressed on the surface of the tumour cells. The method would thus seem limited because a different antibody has to be produced for each individual tumour and because only B-cell tumours, which are uncommon, are known to carry sufficiently specific determinants. Whether it will be possible to find markers that will allow a similar specificity to be gained against other tumours is not yet clear.

Several other patients have, however, already been treated with monoclonal antibodies and it is possible to assess something of the general potential of antibodies for therapy.

The questions to be faced are: are the monoclonal antibodies safe, and are they effective? So far, there is a clear answer to only the first question. Administration of mouse monoclonal antibodies produces remarkably few side effects even when free antigen is present in the serum and it must be presumed that the infusion of antibody leads to the formation of circulating immune complexes[1]. At first sight the apparent safety of monoclonal antibodies might be assumed to stem from their mono-specificity — they do not contain the contaminating specificities that give problems when conventional antibodies, for example anti-lymphocyte globulin, are administered. This may, however, be an over-simplification since it has been predicted that many monoclonal antibodies will give unexpected cross-reactivities[2]. Indeed, cross-reactivity has been demonstrated for some of the antibodies which have been used *in vivo*: for example, the J5 antibody used in treatment of acute leukaemias[3] was found also to react with renal tubules. The moral to be drawn is that before an antibody is used *in vivo* it is wise to screen widely for unexpected reactions with different tissues.

If the antibodies are safe the major problem of antibody specificity remains. Antibodies to differentiation antigens can be used when the loss of normal cells with the same antigens would not matter or when tumour cells have completely replaced the normal population. Thus, antibodies against T-lymphocyte antigens have been used to treat advanced stages of mature T-cell tumours. In the first patient, a significant response of skin lesions was seen and treatment was continued for 17 weeks until the tumour progressed in lymphoid organs[4]. This patient showed no immune response to the mouse immunoglobulin but in subsequent trials and in patients given anti-T cell antibody as immunosuppression for renal allograft rejection episodes, antibody to the mouse protein was produced and limited the duration of treatment[5].

While antibodies to differentiation antigens can be used for tumour therapy, their potential would be much greater if tumour-specific antibodies were available. Several claims for the existence of such antibodies have been made, notably of those to a common acute lymphoblastic leukaemia antigen (CALLA)[6]. However, further examination in another laboratory showed the antigen to be present on some normal bone marrow cells[7], suggesting it might be best to treat all claims with some scepticism, at least until candidate antibodies have been examined in several laboratories. Nevertheless, an anti-CALLA antibody (J5) was used to treat leukaemic patients and revealed another problem of antibody therapy, that the tumour cells rapidly lost the target antigen when antibody was administered. Similar antigenic modulation could be demonstrated *in vitro*.

While treatment *in vivo* has not yet produced spectacular results, with the exception of the anti-idiotype antibody, treatment *in vitro* holds more immediate promise. Bone marrow grafting has become a recognized treatment for leukaemia but has considerable problems, including graft-versus-host disease (GVH) caused by T lymphocytes in the donor marrow. Removal of T lymphocytes from bone marrow before grafting had been attempted with conventional antisera and it was a logical step to see if better results could be obtained with more specific monoclonal antibodies. So far the published data are limited to a single study in which the marrow cells were merely incubated with monoclonal anti-T cell antibody before infusion[8]. The procedure relies for effect on the removal of antibody-coated cells from the circulation by the reticuloendothelial system — hardly likely to be very effective in transplant recipients who have been irradiated and treated with cytotoxic drugs to condition them to accept the graft. Nevertheless, the early results are not discouraging and the use of antibodies with complement or coupled to drugs or toxins should improve its effectiveness. Leukaemia can also be treated by an autologous transplant. Remission bone marrow is collected and stored until the patient relapses. Aggressive chemo- and radiotherapy can then be administered and the patient rescued with the stored marrow. This approach suffers from the problem that the re-infused marrow may contain leukaemic cells but these might be removed if sufficiently specific antibodies can be produced[9]. *In vitro* treatment has the advantage that surplus antibody, or antibody–drug or antibody–toxin conjugate, can readily be washed away before the marrow is infused.

It is notable that the research discussed so far all deals with tumours of the haematopoietic system and common tumours (all carcinomas) have not been mentioned at all. This is in part because many of the earliest and best characterized monoclonal antibodies have been directed to leukocyte antigens, and also because the therapy of solid tumours poses additional problems, particularly the penetration of antibody into tumour masses. Nevertheless

P.C.L. Beverley is in the ICRF Human Tumour Immunology Group, School of Medicine, University College London, University Street, London WC1E 6JJ.

0028-0836/82/220358-02$01.00

an initial trial has been made, although without great effect, in patients with extensive disease[10]. Once again, bone marrow clearance *in vitro* may soon be the most useful manouvre, at least for some solid tumours which are chemo- or radiosensitive.

What of the future? First it is clear that antibody alone is unlikely to be optimally effective and drug- or toxin-coupled antibodies may provide additional therapeutic advantage, particularly *in vitro*. *In vivo* therapy is frequently limited by the development of anti-mouse immunoglobulin antibody but it may be possible to induce tolerance or, in the future, to reduce the problem by the use of human monoclonal antibodies. Already the development of human anti-tumour antibodies has been reported[11]. While serotherapy will surely not be a panacea, there are grounds for cautious optimism.

1. Miller, R.A., Maloney, D.G., Warnke, R. & Levy, R. *New Engl. med. J.* **306**, 517 (1982).
2. Lane, D. & Koprowski, H. *Nature* **296**, 200 (1982).
3. Ritz, J. *et al. Blood* **58**, 141 (1981).
4. Miller, R.A., Maloney, D.G., McKillop, J. & Levy, R. *Blood* **58**, 78 (1981).
5. Cosimi, A.B. *et al. New Engl. med. J.* **305**, 308 (1981).
6. Ritz, J. *et al. Nature* **283**, 583 (1980).
7. Greaves, M.F. *et al.* in *Modern Trends in Human Leukaemia* Vol. 4 (eds Neth, R. *et al.*) 296 (Springer, Berlin, 1981).
8. Prentice, H.G. *et al. Lancet* i, 700 (1982).
9. Storb, R. *Nature* **295**, 555 (1982).
10. Sears, H.F. *et al. Lancet* i, 762 (1982).
11. Sikora, K. *et al. Lancet* i, 11 (1982).

Plant cells on the move

from Clive Lloyd

CORSETED in cellulose, plant cells have limited freedom of movement. As a result, their morphogenetic responses, rather than consisting of cellular migration, involve internal rearrangements of cytoplasm. The responses include alterations in the rates of cell division and of enlargement (which determine the overall size of cells). Fine tuning of tissue morphology is, however, largely a product of the shapes adopted by cells and of the orientation of the walls which are deposited between them at cytokinesis. A common element uniting all these processes is the microtubular cytoskeleton since microtubules: (1) help to determine the polarized shape of interphase cells by somehow guiding the deposition of cellulose; (2) as the preprophase band (PPB) they foretell where the cytokinetic cell plate will be deposited; (3) help form the mitotic spindle; and (4) contribute to the cytokinetic apparatus, the phragmoplast. The microtubule cycle can therefore be regarded as at the hub of key morphogenetic processes. To come to terms with this at the level of control it will be important to obtain a molecular description to complement the existing ultrastructural one. For instance, where are the microtubule-organizing centres located? Is there one set responsible for all microtubule arrays or are there several? If several, how are their functions coordinated?

In comparison with animal tissue culture cells, plant cells are much less tractable and, as a consequence, the successful application of molecular probes to the study of the cytoskeleton has been slow. But over the last year several studies have emerged which present real technical advances that should lead to a more dynamic picture of the microtubule cycle.

One possibility when considering the microtubule cycle is that different tubule arrays are composed of different tubulins produced, perhaps, in a cycle-dependent manner. The multitubulin concept was reviewed by Fulton in a recent *News and Views* article[1] which outlined that non-mammalian organisms do contain multiple genes for tubulins and that the α-subunits of their tubulin heterodimers may differ from those of vertebrate brains, which, unlike most other tissues and unicellular organisms, constitute a rich source of microtubule proteins. Until now the lack of grey matter has held back research into plant tubulin, but in this issue of *Nature* (p.426), Morejohn and Fosket report its isolation from Paul's scarlet rose suspensions and, for the first time, its reassembly into microtubules.

Plant extracts, prepared according to methods for brain tubulin and containing leupeptin as a protease inhibitor, were fractionated on DEAE-Sephadex. One fraction was enriched in proteins which migrated on SDS gels in the vicinity of brain tubulins. Indeed, microtubules could be produced from this fraction either by self-assembly in glycerol or by 'precipitation' with the plant metabolite taxol which enhances the polymerization of tubulin[2]. It is of interest that comparative peptide maps suggest that whereas brain and plant β-tubulins are very similar, the α-subunits are different. As Clayton *et al.* have suggested[3], this could be the basis of the relative resistance of the microtubules of microorganisms to drugs such as colchicine and this explanation would now seem to apply to higher plants as well. Provided that satisfactory cell synchrony can be obtained, the stage is now set to see whether there are cycle-specific classes of plant tubulins.

Clive Lloyd is at the Unilever Research Colworth Laboratory, Sharnbrook, Bedford MK44 1LQ.

Several recent histological studies have outlined the dynamic nature of plant microtubules. De Mey *et al.*[4] reported the staining of microtubules in the endosperm of *Haemanthus* (giant African blood lily) by an immuno-gold method. Using this technique, anti-tubulin antibody-decorated microtubules can be made visible by the subsequent addition of colloidal gold-tagged antibodies and beautiful pictures have been obtained of interphase and mitotic arrays. One point to emerge from this work is that cytoplasmic microtubules seem to be grouped around the nucleus to the extent that it is concluded "that the nuclear envelope or the nucleus, or both, is instrumental in tubule nucleation and arrangement". This is intriguing since sites of microtubule nucleation in acentriolar plants are poorly understood and difficult to pinpoint, and this study gives (some of) them a definite perinuclear location. The function of these cytoplasmic microtubules in endosperm (which, at stages, is liquid; containing wall-less nuclei in a syncitium) is enigmatic and had not been previously detected. It is therefore possible that they have a different function to the plasma membrane-associated microtubules believed to help orientate cellulose microfibrils and which have been reported to emerge, in *Azolla* roots, from nucleation sites located at the cell periphery[5]. Another interesting feature of this study is that it shows cytoplasmic microtubules co-existing with phragmoplast tubules which, in turn, overlap in time with the anaphase mitotic spindle. This implies separate microtubule-organizing centres or nucleation events for different microtubule assemblies whose bouts of activity are not mutually exclusive.

The large nuclei of *Haemanthus* endosperm make it ideal material for the study of mitosis and, as Euteneur and McIntosh[6] have shown with their work on microtubule polarity, for phragmoplast formation also. Interphase cytoplasmic hoops of microtubules are absent from this tissue, however, and other material must be used to study them. The difficulty with applying antibody-based techniques to walled cells is that the wall generally excludes antibodies, but there are now ways around this problem. By pre-fixing roots with aldehydes, Wick *et al.* [7] were able to preserve cell shape before degrading the wall with enzymes and with EGTA (which presumably weakened intercellular calcium pectinates). Following this treatment, gentle squashing released cuboidal cells which could at this stage be permeabilized with acetone before carrying out double immunofluorescence using anti-tubulin antibodies. For the first time, the PPB has been visualized by immunofluorescence and all four assemblies are beautifully illustrated. Some interesting aspects deserve comment. For example, at the same time as the PPB is stained, so is the nuclear envelope, which appears to have

0028-0836/82/220359-02$01.00

microtubules — not of the PPB — associated with it. In some cases where the nucleus has been freed from the rest of the cell, the PPB segregates with the nucleus and not with the cell cortex.

It is early days in this more detailed phase of plant cell biology but already the new approaches are presenting tantalizing and unexpected findings. With such techniques now successfully applied to higher plants, we can expect to learn more about the spatio-temporal aspects of microtubule nucleation which accompany key processes in plant morphogenesis. □

1. Fulton *Nature* **296**, 308 (1982).
2. Schiff *et al. Nature* **277**, 665 (1979).
3. Clayton *et al. FEBS Lett.* **115**, 301 (1980).
4. De Mey *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 1898 (1982).
5. Gunning *et al. Planta* **143**, 161 (1978).
6. Euteneur & McIntosh *J. Cell Biol.* **87**, 509 (1980).
7. Wick *et al. J. Cell Biol.* **89**, 685 (1981).

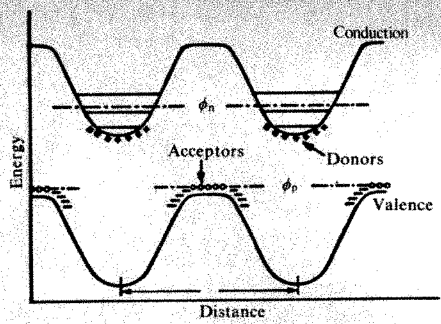


Fig. 2 Spatial variation of the conduction and valence bands in an excited state of the superlattice. The extra electrons (gathered in conduction valleys) and holes (gathered near peaks in valence band) are implied by the separation of chemical potentials, ϕ_n and ϕ_p, for the two species. Recombination is slowed by the spatial separation of electrons from holes. After ref. 1.

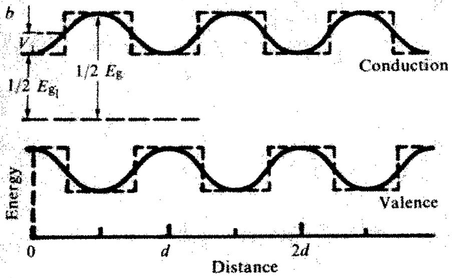
Electrons in novel two-dimensional structures

from J.M. Worlock

THE report that Gottfried H. Döhler and his colleagues have been able to grow a specimen of gallium arsenide (GaAs) in the form of a doping superlattice[1] is an exciting landmark in semiconductor physics. For Döhler, primarily a theoretician, at the Max-Planck Institut fur Festkorperforschung in Stuttgart, the development is the culmination of ten years' work. Having discerned the properties expected of a doping superlattice[2], he has now persuaded colleagues in Stuttgart and Munich to grow the new lattice and to confirm his predictions.

Interest in semiconductor superlattices goes back to 1970 when Esaki and Tsu[3], at IBM, perceived that it was possible, at least in principle, to improve on nature and design semiconductor microstructures with carefully tailored properties.

Fig. 1 Spatial variation of the conduction and valence bands in two types of semiconductor superlattice: *a*, doping superlattice of alternating n-type and p-type layers; *b*, compositional superlattice with alternation of crystal composition. The solid and dashed lines represent simply sinusoidal and sharp-edged band variation respectively. E_gs are bandgap energies of the constituents and V_1 is the amplitude of the superlattice potential, fixed by the doping levels in *a*, and depending on both the E_gs and electron affinities in *b*, where the amplitudes of variation in valence and conduction bands need not be equal. After ref. 3.



They envisioned two types of semiconductor superlattice. One, a doping superlattice, is that analysed by Döhler (and independently by Yu. A. Romanov[4]) and recently realized. The other, a compositional superlattice, was realized as early as 1974 by groups at IBM[5] and Bell Laboratories[6] and has been much studied[7].

Figure 1 shows the spatial variation of the electron bands in the direction of the superlattice. The doping superlattice (Fig. 1*a*) is seen to those familiar with semiconductors to be simply an alternation of p–n and n–p junctions. The periodic rise and fall of the conduction and valence bands is caused by a periodic variation of impurity concentration. The rise and fall also mimics on a more macroscopic scale the periodic attraction and repulsion that electrons experience in the microscopic atomic lattice: electrons are attracted to minima in the conduction band while holes are attracted to maxima in the valence band.

In a compositional superlattice (Fig. 1*b*), on the other hand, alternating layers are of different elemental composition. Locally, each composition has its own band structure, so that within limits the periodic variation of the conduction and valence bands can be arbitrarily chosen. The periodic structure shown in the figure resembles the familiar case of the combination of GaAs, with bandgap E_{g1}, with an alloy semiconductor aluminium–gallium arsenide (AlGaAs) having a slightly larger bandgap. A more recent development, the type 2 superlattice, conceived at the IBM laboratories[8], is exemplified by the combination of indium arsenide (InAs) and gallium antimonide (GaSb). Here the valence and conduction bands interpenetrate, resulting in a much more complicated superlattice.

Döhler *et al.*'s doping superlattice has been produced by growing a crystal of GaAs by the technique of molecular beam epitaxy (MBE). Planar layers of ~40 nm thickness (~150 atomic layers) were doped alternately with silicon (n-type) and beryllium (p-type) atoms, at densities of about 10^{18} atoms per cm^3. The resulting structure of the superlattice crystal is shown (Fig. 2) in the excited state resulting from illumination, which puts extra holes in the valence band and extra electrons in the conduction band. The extra electrons and holes are spatially separated with the interesting consequence that they can have anomalously long lifetimes, their mutual annihilation or recombination is slowed by their inability to approach each other. A second consequence is that some of the bending of the bands is undone by the extra carriers, leading to a crystal with a 'bandgap' uniquely sensitive to illumination.

Dohler *et al.* have produced unambiguous evidence for this sensitivity (Fig. 3). The quantum energy of the photoluminescence measures this bandgap, since the luminescence is annihilation radiation. At low excitation intensity, the bands shown in Fig. 2 are strongly bent and the gap energy is small, around 1.3 eV; but as the excitation intensity is increased, the gap energy shifts higher, approaching the energy of the bandgap of undoped GaAs, at about 1.53 eV.

Further evidence for the formation of a

J.M. Worlock is in the Bell Telephone Laboratories, Holmdel, New Jersey 07733. A. Pinczuk collaborated in the preparation of the article, and D. Olego provided helpful criticism.

1. Döhler, G.H. *et al. Phys. Rev. Lett.* **47**, 864 (1981).
2. Döhler, G.H. *Phys. Status Solidi* **52**, 79, 533 (1971); *J. Vac. Sci. Technol.* **16**, 851 (1979).
3. Esaki, L. & Tsu, R. *IBM J. Res. Devel.* **14**, 61 (1970).
4. Romanov, Yu. A. *Sov. Phys. Semiconductors* **5**, 1256 (1972); Romanov, Yu. A. & Orlov, L.K. *Sov. Phys. Semiconductors* **7**, 182 (1973).
5. Ludeke, R., Esaki, L. & Chang, L.L. *Appl. Phys. Lett.* **24**, 417 (1974).
6. Dingle, R., Wiegmann, W. & Henry, C.H. *Phys. Rev. Lett.* **33**, 827 (1974).
7. For a recent review, see Gossard in *Thin Films: Preparation and Properties* (eds Tu, K.N. & Rosenberg, R.) (Academic, New York, 1981).
8. Sai-Halasz, G.A., Tsu, R. & Esaki, L. *Appl. Phys. Lett.* **30**, 651 (1977).
9. Dingle, R. *Adv. Solid State Phys.* **15**, 21 (1975).
10. Klitzing, K.V., Dorda, G. & Pepper, M. *Phys. Rev. Lett.* **45**, 494 (1980).
11. Tsui, D.C., Gossard, A.C., Feld, B.F., Cage, M.E. & Dziuba, R.F. *Phys. Rev. Lett.* **48**, 3 (1982).

0028-0836/82/220360-02$01.00

high-quality doping superlattice is given in the form of inelastic light-scattering spectra, which show how the electrons are distributed in their valleys, and how these distributions can be changed by optical excitation.

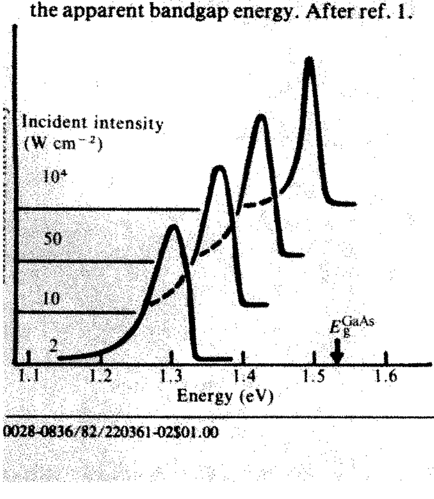
The techniques used in growing semiconductor superlattices are potentially of great importance. MBE, which originated at Bell Laboratories in the late 1960s, is one of a class of epitaxial crystal growth processes in which new material forms in thin layers on top of and in register with an existing crystalline substrate. These processes differ from each other in the manner in which new atoms are brought to the growth surface. Although other epitaxial processes have grown excellent crystals, precise control down to monatomic layers is still attained only with the MBE technique. Of course, it is not only superlattices which are grown: many of the most important structures involve simply one or two interfaces or heterojunctions between dissimilar semiconductors.

Among the important optical devices that have been fabricated starting with epitaxial growth are light-emitting diodes (LEDs), lasers and detectors; all of these are useful in the expanding field of optical communications, while LEDs also have important applications as display devices.

Because of the precise control which is possible in MBE growth and because of the invention, at Bell Laboratories, of modulation doping, researchers have been able to construct extremely smooth planar (two-dimensional) pathways for electronic conduction. Highly mobile electrons travelling in these layers may be involved in tomorrow's high-frequency electron devices.

But equally important is the 'pure physics' excitement which has attended the refinement of the MBE technique. Some of the earliest experiments were used to demonstrate the quantum-mechanical particle-in-a-box behaviour of electrons and Wannier excitons in AlGaAs–GaAs superlattices[9]. More recently some

Fig. 3 Optical spectra of the recombination luminescence emitted from a doping superlattice, showing how the increasing incident intensity shifts the radiation to higher frequency, implying that larger populations of electrons and holes change the apparent bandgap energy. After ref. 1.



unexpected peculiarities of matter in two-dimensional space (2D) have received attention. The effects of randomness or roughness of the pathway have a powerful influence on the localization of mobile particles in 2D, and studies of electron localization in epitaxial layers have increased our understanding of this concept.

A startling effect in 2D transport was discovered in 1980 by Klaus von Klitzing and his co-workers[10]. Their anomalous Hall effect has been developed in MBE-grown AlGaAs–GaAs heterostructures to give a new precision to the value of the fine structure constant[11], and has also led to deeper theoretical insight into matter of reduced dimensionality.

Many more examples of exciting developments could be adduced, showing that new structures can be exploited both for new devices and for new science. I expect the doping superlattice to follow in this honourable tradition. □

Isotopic anomalies in meteorites

from Edward R.D. Scott

WHAT D.D. Clayton[1] has called a new field of astronomy, 'astrochemistry', is developing from an appreciation[2] that the Solar System did not form from a hot well-mixed cloud of uniform chemical and isotopic composition, as was once believed. Exciting discoveries of isotopic anomalies in meteorites, made during the past decade, are being used to test and develop theories for the formation of elements in stars, the formation and history of interstellar dust and the origin of the Solar System. Studies of the isotopic composition of oxygen and magnesium in chondrites provide crucial evidence for a dust component in the early Solar System containing essentially pure ^{16}O, and ^{26}Mg excesses that result from the decay of ^{26}Al. Relatively simple, attractive and testable models have been proposed to explain the oxygen isotopic abundances, but the interpretation of the ^{26}Mg excesses is more controversial.

R.N. Clayton and his associates[3–6] have established that the oxygen isotopic compositions of minerals and components in chondrites are a result of various degrees of mixing between reservoirs with intrinsically different isotopic compositions. On a plot of $^{17}O/^{16}O$ against $^{18}O/^{16}O$, data for chondrules and nearly all Ca–Al-rich inclusions define a set of closely spaced lines with slopes of 0.94–0.97. (Terrestrial samples define a single line with a slope of 0.52, because of mass-dependent fractionation processes.) Three lines have so far been identified for H-L-LL, CM and CO-CV chondrite groups, but others may be resolved. On the three-isotope plot, the lines appear to radiate from a point marking the oxygen composition of spinels, which occur in Ca–Al-rich inclusions and have the lowest $^{17}O/^{16}O$ and $^{18}O/^{16}O$ ratios of all minerals. Isotopic compositions of minerals may also depend on the nature of the host chondrule or inclusion; the largest variations are found in the Ca–Al-rich inclusions of carbonaceous chondrites.

Oxygen isotopic data suggest that meteorite components largely formed from two ingredients: interstellar gas with a slightly variable isotopic composition, and well-mixed interstellar dust, about 5 per cent of which contained essentially pure ^{16}O (refs 3,7,8). Dust, or chondrules and inclusions derived from this dust, exchanged oxygen with the gas to various degrees. Minerals such as spinel are presumed to preserve the original isotopic composition of the dust because of their low oxygen diffusion rates. (Oxygen diffusion rates in these minerals have not yet been measured.) FUN inclusions, which are those with many large isotopic anomalies[2], may have formed from poorly mixed interstellar dust balls[8]. During their formation, these inclusions experienced large isotopic fractionation, before partial exchange of oxygen with the gas[9]. Oxygen shows the most widespread isotopic anomalies because it was the only element that was abundant in both nebula gas and dust[8].

Wood[8] deduces from these and other data that the Ca–Al-rich inclusions and chondrules did not form by condensation from a gaseous solar nebula[10], as many once believed. He supports D.D. Clayton's view[11] that they formed from unvaporized aggregates of interstellar grains. Ca–Al-rich inclusions probably formed either as distillation residues during unspecified heating events in the solar nebula[8], or as condensates from supernova ejecta[11].

In contrast, the picture obtained from studies of magnesium isotopic abundances is relatively confusing. Excesses of ^{26}Mg formed by decay of ^{26}Al have been found in several minerals with high Al/Mg ratios[12]. About 15 Ca–Al-rich inclusions in the Allende meteorite have been found to contain ^{26}Mg excesses that are correlated with the Al/Mg ratio. However, the proportion of their Al atoms that were radioactive when the inclusions formed (the initial $^{26}Al/^{27}Al$ ratio) varied widely from $< 2 \times 10^{-7}$ to 10^{-3} (see ref. 12). Some types of inclusion (type B1) have values of

Edward R.D. Scott is at the Department of Geology and Institute of Meteoritics, University of New Mexico, Albuquerque, New Mexico 87131.

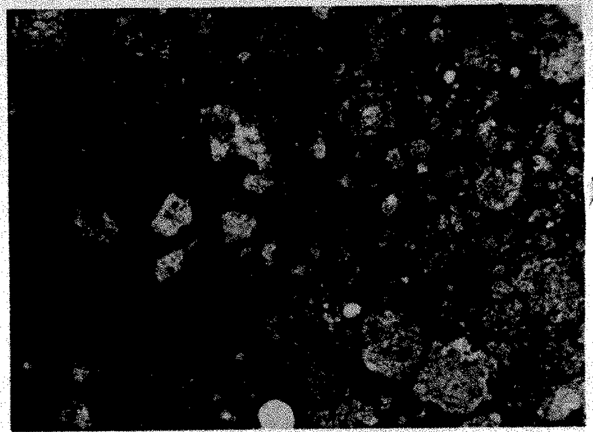
0028-0836/82/220361-02$01.00

5×10^{-5} for this ratio, but in other inclusions, various crystals or minerals may give diverse values[13]. These differences may be due to a difference of up to 9 Myr in their times of formation (or final internal equilibration). Alternatively, $^{26}Al/^{27}Al$ ratios may have varied widely in the solar nebula[12]. However, the size of isotopic variations that would be needed to account for simultaneous formation of these inclusions is much larger than has been found for other elements.

Bar-Matthews *et al.*[14] have recently analysed an unusual Ca–Al-rich inclusion, which contains corundum (Al_2O_3), in the Murchison meteorite. If the inclusion had formed by equilibrium condensation from a slowly cooling gas of solar composition (which now seems rather unlikely), corundum should contain a large ^{26}Mg excess, as it is the first major mineral predicted to condense[10] and has a very high Al/Mg ratio. Instead, the authors found only small ^{26}Mg excesses, which were uncorrelated with the Al/Mg ratio and which they attributed to Mg isotopic heterogeneities.

The lack of excess ^{26}Mg in Murchison corundum and the wide variation in the initial $^{26}Al/^{27}Al$ ratios of Allende inclusions are cited by D.D. Clayton[1,15] as evidence against decay of ^{26}Al in the Solar System. He rejects explanations of time differences or isotopic heterogeneities and argues that ^{26}Al decayed in interstellar grains. He attributes correlated ^{26}Mg excesses in individual anorthite crystals to precursor Al_2O_3 stardust. This controversial idea has stimulated more precise experiments, but is still a minority view.

Analogous problems have arisen in the interpretation of ^{129}Xe excesses in ordinary chondrites due to decay of ^{129}I, which has a half life of 17 Myr. If ^{129}I was homogeneously distributed in the solar nebula, then two chondrules in the Bjurbole chondrite formed 2 Myr apart[16], a surprisingly long time. D.D. Clayton has also questioned the existence of live ^{129}I in the Solar System[1], but the best working hypothesis seems to be that the I–Xe time differences, like those for the Al–Mg chronometer, may be real. Some authors have invoked planetary, rather than nebular, processes to form chondrules[17,18] and Ca–Al-rich inclusions[19]. This might help to explain the large spread of formation ages of chondrules and inclusions, but the oxygen isotopic data strongly suggest that chondrules and Ca–Al-rich inclusions are all primitive objects, which may have formed during similar high-temperature events[20].



Transmitted light photograph of the Vigarano carbonaceous chondrite showing Ca-Al-rich inclusion with complex rims and smaller chondrules. Actual size: 8 × 11 mm.

Progress in astrochemistry will be greatly advanced when the origin of the chondrules and Ca–Al-rich inclusions is better understood. This probably requires a realization that all the inhabitants in the exotic menagerie of chondritic components — Ca–Al-rich inclusions and chondrules, mafic chondrules, aggregates and matrix, and metallic Fe, Ni — have interrelated origins. The origin of chondrules in enstatite chondrites, for example, cannot be divorced from an understanding of the formation of Ca–Al-rich inclusions in carbonaceous chondrites.

1. Clayton, D.D. *Q. Jl R. astr. Soc.* (in the press).
2. Begemann, F. *Rep. Prog. Phys.* **43**, 1309–1359 (1980).
3. Clayton, R.N., Onuma, N., Grossman, L. & Mayeda, T.K. *Earth planet. Sci. Lett.* **34**, 209–224 (1977).
4. Gooding, J.L. *et al. Meteoritics* **15**, 295 (1980).
5. Clayton, R.N., Mayeda, T.K., Gooding, J.L., Keil, K. & Olsen E.J. *Lunar planet. Sci.* **12**, 154–156 (1981).
6. Clayton, R.N. & Mayeda, T.K. *Lunar planet. Sci.* **13**, 117–118 (1982).
7. Blander, M. & Fuchs, L.H. *Geochim. cosmochim. Acta* **39**, 1605–1620 (1975).
8. Wood, J.A. *Earth planet. Sci. Lett.* **56**, 32–44 (1981).
9. Clayton, R.N. & Mayeda, T.K. *Geophys. Res. Lett.* **4**, 295–298 (1977).
10 GRossman, L. *A. Rev. Earth planet. Sci.* **8**, 559–608 (1980).
11. Clayton, D.D. *Astrophys. J.* **199**, 765–769 (1975); *Moon Planets* **19**, 107–137 (1978).
12. Lee, T. *Rev. Geophys. Space Phys.* **17**, 1591–1611 (1979).
13. Hutcheon, I.D. *Am. chem. Soc. Symp. Ser.* **176**, 95–128 (1982).
14. Bar-Matthews, M., Hutcheon, I.D., MacPherson, G.J. & Grossman, L. *Geochim. cosmochim. Acta* **46**, 31–41 (1982).
15. Clayton, D.D. *Lunar planet. Sci.* **13**, 115–116 (1982).
16. Caffee, M.W., Hohenberg, C.M., Hudson, B. & Swindle, T.D. *Lunar planet. Sci.* **13**, 75–76 (1982).
17. Leitch, C.A. & Smith, J.V. *Nature* **290**, 578–579 (1980).
18. Zook, H. *Lunar planet. Sci.* **12**, 1242–1244 (1981).
19. Armstrong, J.T., Meeker, G.P., Huneke, J.C. & Wasserburg, G.J. *Geochim. cosmochim. Acta* **46**, 575–595 (1982).
20. Scott, E.R.D., Taylor, G.J. & Keil, K. *Lunar planet. Sci.* **13**, 704–705 (1982).

A new model for nitrogen control

from M.J. Merrick

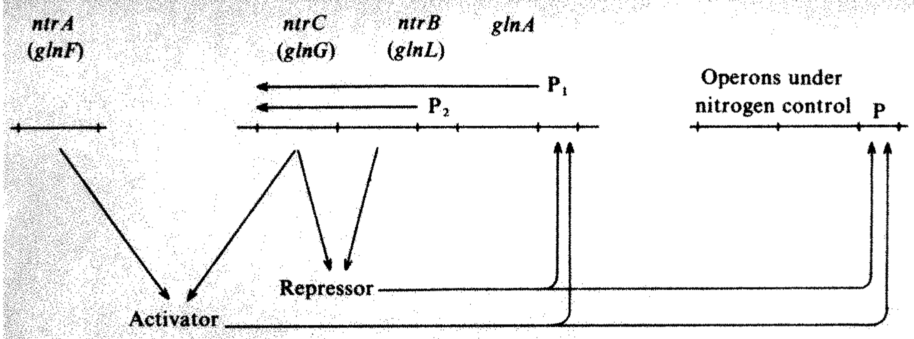
IN *Escherichia coli* and other enteric bacteria, the induction of a number of operons involved in carbon assimilation, for example, *lac, ara, gal,* is affected by the intracellular concentration of cyclic AMP, which is in turn determined by the carbon source available to the cell. This phenomenon is termed catabolite repression and is mediated by a complex of cyclic AMP and the cyclic AMP receptor protein (CRP) which together form a transcriptional activator. There has long been evidence[1] for another regulatory system controlling the balance of nitrogen assimilation enzymes in bacteria, for example, histidine (*hut*)and proline (*put*) utilization, and nitrogen fixation (*nif*). This phenomenon, which has been termed 'nitrogen control'[2], has received considerably less attention than catabolite repression although a model for control of nitrogen assimilation was proposed in 1973 (refs 3,4). Recent evidence from genetic studies in a number of organisms suggests that the system is considerably more complex than originally proposed and involves both positive and negative regulatory elements.

Studies on the regulation of histidase synthesis in *Klebsiella aerogenes* originally led Magasanik and co-workers to formulate a model[5] in which the principal regulator of nitrogen assimilation was the enzyme glutamine synthetase (GS). They postulated that GS was responsible not only for the biosynthesis of glutamine and the assimilation of nitrogen into glutamate, but also for the activation of transcription of operons such as *hut* and *put*. The enzymatic activity of GS is regulated by the reversible adenylylation of specific tyrosyl residues on each subunit of this dodecameric protein. In conditions of nitrogen starvation, GS is deadenylylated and highly active, and Magasanik *et al.* proposed that this deadenylylated form of GS was an activator of transcription. Evidence for the model included the isolation of mutants, believed to be in the GS structural gene (*glnA*), which showed altered regulatory properties[5].

In the last five years, studies of nitrogen control have been extended to *Salmonella typhimurium, E. coli* and *Klebsiella pneumoniae* as well as *K. aerogenes.* A very different model for regulation of nitrogen assimilation has now emerged in which GS is no longer considered to be a regulatory protein and instead three specific genes, *ntrA, ntrB* and *ntrC* (or their equivalents), have been identified in all the above four species. Two of these genes, *ntrB* and *ntrC,*

M.J. Merrick is in the Agricultural Research Council Unit of Nitrogen Fixation, University of Sussex, Brighton BN1 9RQ.

0028-0836/82/220362-02$01.00



Model for nitrogen control in *S. typhimurium* and *E.coli*.

are contiguous with the GS structural gene, *glnA*, and this close linkage confounded the earlier analysis of mutants.

The *ntrA* gene (originally designated *glnF*) was first identified in *S. typhimurium*[6]. *NtrA* is not linked to *glnA* but mutations in this gene result in constitutive low levels of GS and glutamine auxotrophy[6–8]. These mutants are also unable to activate other genes under nitrogen control.

Following identification of *ntrA*, another regulatory locus closely linked to *glnA* was described in *S. typhimurium*[2], *E. coli*[9] and *K. pneumoniae*[8,10,11]. Then at the 1981 meeting of the American Society for Microbiology in Dallas, it was reported that in both *S. typhimurium* (Kustu and colleagues at the University of California, Davis) and *E. coli* (Tyler and colleagues at the Massachusetts Institute of Technology), this regulatory locus consists of two cistrons (*ntrB* and *ntrC* in *S. typhimurium*, and *glnL* and *glnG* in *E. coli*). Tyler *et al.* have shown that in *E. coli*, the two genes constitute a single operon transcribed from *glnL* to *glnG* and it has been proposed that this operon may be transcribed either from its own weak promoter (P_2) or by stronger read-through transcription from the *glnA* promoter (P_1)[12]. The products of the *E. coli ntrB* and *ntrC* genes have been identified as polypeptides of molecular weights 36,000 (*ntrB*)[14] and 54,000 (*ntrC*)[13,14].

The glutamine auxotrophy induced by *ntrA* mutations can be suppressed by mutations in either *ntrB* or *ntrC* with a resultant constitutive low level of GS synthesis. Hence it seems that in *ntrA* mutants, GS is permanently repressed and this repression is relieved by *ntrB* or *ntrC* mutations, leaving a low level of unregulated *glnA* transcription from P_1. This suppression does not, however, extend to other operons under nitrogen control which fail to be activated in all *ntrA*$^-$ strains. In an *ntrA*$^+$ background, *ntrB* mutations result in constitutive expression of *glnA* whereas most *ntrC* mutations prevent expression of *glnA* and other genes under nitrogen control.

From the extensive genetic analysis in both *S. typhimurium* and *E. coli* a consensus model (see the figure) has emerged for nitrogen control in the two organisms. It is proposed that the *ntrB* and *ntrC* products can repress or activate transcription of genes under nitrogen control, including *glnA*. Repression requires both the *ntrB* and *ntrC* products (possibly acting as a complex), hence the suppression of the *ntrA* Gln$^-$ phenotype by mutations in either *ntrB* or *ntrC*. However, activation apparently requires only the *ntrC* (and *ntrA*) product(s) because *ntrB* mutants express *glnA* constitutively. The Gln$^-$ phenotype of certain *ntrC* mutants is proposed to be due to the *ntrB–ntrC* complex being permanently locked in the repressor form, a model which is supported by the isolation of Gln$^+$ suppressors of such *ntrC* mutations in both *ntrB* and *ntrC*[15].

The *ntrC* product therefore seems likely to be a DNA-binding protein which can activate transcription. The *ntrB* product could also bind to DNA in the manner of a conventional repressor or alternatively could modify the activator properties of the *ntrC* product. The *ntrA* product is a positive regulatory factor necessary for the formation of a functional *ntrC* activator and hence *ntrA* mutants cannot derepress genes under nitrogen control. Analogies between this system and the cyclic AMP–CRP system for carbon and energy regulation are striking and extension of such analogies suggests that *ntrA* may encode an enzyme that synthesizes a low molecular weight co-regulator of transcription — a signal of nitrogen deficiency[2]. It is of interest that certain operons, for example, *hut*, are subject to regulation by catabolite repression and by nitrogen control mediated by the elements described above.

The *ntrB*, *ntrC* regulatory system in which two genes in a single operon encode a repressor and an activator respectively is almost unique among the bacterial control systems so far analysed. However, one other system shows striking analogies to that proposed for *ntrB* and *ntrC*.

The nitrogen fixation (*nif*) regulon in *K. pneumoniae* is controlled by two *nif*-specific regulatory genes, *nifL* and *nifA*, which act respectively to repress or activate *nif* transcription[16,17]. These two genes comprise a single operon, *nifLA*, and their products may also function as a protein complex. As with *ntrB,C*, the product of one *nif* gene, *nifA*, is alone sufficient for activation of transcription[18]. Interestingly the *nifLA* operon is in turn regulated by the general nitrogen control system which our recent studies have shown to be essentially identical to that described for *ntrA,B,C* in *E. coli* and *S. typhimurium*.

The recent work described above should stimulate new interest in the regulation of nitrogen assimilation, but just as the mechanism of catabolite repression is still not completely understood[19], the details of nitrogen control will probably take considerable time to dissect. In the meantime, a recent report of a regulatory locus closely linked to the structural gene for GS in *Bacillus subtilis*[20] prompts speculation as to whether nitrogen control in organisms other than the enteric coliform bacteria is mediated by a similar genetic system. □

1. Neidhart, F.C. & Magasanik, B. *J. Bact.* **73**, 253 (1957).
2. Kustu, S., Burton, D., Garcia, E., McCarter, L. & McFarland, N. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4576 (1979).
3. Brenchley, J.E., Prival, M.J. & Magasanik, B. *J. biol. Chem.* **248**, 6122 (1973).
4. Prival, M.J., Brenchley, J.E. & Magasanik, B. *J. biol. Chem.* **248**, 4334 (1973).
5. Magasanik, B. *et al. Curr. Topics Cell Regulation* **8**, 119 (1974).
6. Garcia, F., Bancroft, S., Rhee, S.G. & Kustu, S. *Proc. natn. Acad. Sci. U.S.A.* **74**, 1662 (1977).
7. Gaillardin, C.M. & Magasanik, B. *J. Bact.* **133**, 1329 (1978).
8. Leonardo, J.M. & Goldberg, R.B. *J. Bact.* **142**, 99 (1980).
9. Pahel, G. & Tyler, B. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4544 (1979).
10. de Bruijn, F.J. & Ausubel, F.M. *Molec. gen. Genet.* **183**, 289 (1981).
11. Espin, G., Alvarez-Morales, A. & Merrick, M. *Molec. gen. Genet.* **184**, 213 (1981).
12. Magasanik, B. & Rothstein, D.M. in *Glutamine: Metabolism, Enzymology and Regulation* (eds Mora, J. & Palacios, R.) 61 (Academic, New York, 1980).
13. Backman, K., Chen, Y-M. & Magasanik, B. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3743 (1981).
14. McFarland, N., McCarter, L., Artz, S. & Kustu, S. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2135 (1981).
15. Wei, G.R. & Kustu, S. *Molec. gen. Genet.* **183**, 392 (1981).
16. Dixon, R. *et al. Nature* **286**, 128 (1980).
17. Merrick, M. *et al. Molec. gen. Genet.* (in the press).
18. Buchanan-Wollaston, V., Cannon, M.C., Beynon, J. & Cannon, F.C. *Nature* **294**, 776 (1981).
19. Ullmann, A. & Danchin, A. *Trends biochem. Sci.* **5**, 95 (1980).
20. Reysset, G. *J. Bact.* **148**, 653 (1981).

100 years ago

The Photographic Gun — It takes twelve images, in one second, of an object on which the piece is continuously sited. The exposure time of each image is 1-720th of a second. It is being applied to the taking of birds flying.

From *Nature* **26**, 25 May 1882.

0028-0836/82/220363-01$01.00

Vitamin D: sunlight and precursors

from Alastair Hay

WITH some twenty metabolites of vitamin D now identified (see the figure for some of them) chemists in the field have much to be satisfied about. There is less satisfaction, however, for those concerned with the biological importance of these metabolites and a recent workshop* on vitamin D raised more questions than it answered.

Take, for example, the role of the two kidney metabolites of vitamin D, 1,25-dihydroxyvitamin D (1,25-$(OH)_2D$) and 24,25-dihydroxyvitamin D (24,25-$(OH)_2D$). The former is the active hormonal form of vitamin D. The controversial suggestion that 24,25-$(OH)_2D$ is also a potential hormone active in bone was made several years ago but it is still far from clear whether 1,25-$(OH)_2D$ alone is necessary to cure the bone disease osteomalacia — caused by a deficiency of vitamin D — or whether both metabolites are required.

Obtaining an adequate supply of vitamin D is a greater problem for old people than for the young. The two main sources are from food and exposure to UV light in the summer. For most people in Britain, including the elderly, it is the latter source which is the most important (Lawson *et al. Br. med. J.* **II**, 303; 1979).

Although both young and old are said to produce equivalent amounts of vitamin D when exposed to UV light (Davie & Lawson *Clin. Sci.* **58**, 235; 1980), the concentration of the precursors of vitamin D_3 in the skin seems to fall with age. M.F. Holick (Massachusetts General Hospital) reported that the epidermal stores of 7-dehydrocholesterol (7-DHC) are inversely related to age, as is the amount of pre-vitamin D_3 (pre-D_3) formed from 7-DHC after UV irradiation.

One possible regulator of 7-DHC stores is the kidney metabolite 1,25-$(OH)_2D$. Holick has detected a high-affinity, low-capacity receptor for 1,25-$(OH)_2D_3$ in fibroblasts and keratinocytes cultured from skin biopsies taken from normal volunteers that was absent in cells cultured from a skin biopsy of a patient with vitamin D-dependent rickets type II. The kidney metabolite increased the ratio of 7-DHC to cholesterol in cell cultures of keratinocytes and Holick suggested that stores of 7-DHC in skin may be controlled by 1,25-$(OH)_2D_3$ inhibiting the conversion of 7-DHC to cholesterol.

The steps in the conversion of 7-DHC to pre-D_3 and then vitamin D_3 are controlled by UV light and temperature respectively. The vitamin is removed from the skin attached to a specific vitamin D-binding protein (DBP) which has a much higher affinity for vitamin D_3 than for pre-D_3 and so allows only the former to enter the circulation.

There has been much speculation about the function of DBP and argument about the importance of free and bound vitamin D_3 metabolites in plasma. It now seems clear that the major role of DBP is as a storage protein for vitamin D metabolites, for which it has a high affinity.

Recently Haddad *et al.* (*J. clin. Invest.* **67**, 1550; 1981) suggested that vitamin D and its metabolites attached to DBP might first bind to a membrane receptor and then be pinocytosed into cells. R. Bouillon (University of Leuven) reported, however, that he could find no sign of a membrane receptor for DBP. Labelled DBP did not bind to kidney cells in culture in any appreciable quantity and the little that did bind was not altered by the presence of either vitamin D metabolites or unlabelled DBP. The binding was therefore not specific, a finding which Bouillon confirmed using immunofluorescent techniques.

Thus Bouillon concluded that DBP acts like other transport proteins for hormones and that it is the free hormone which crosses cell membranes. In normal subjects, a positive correlation was found between the serum concentrations of DBP and 1,25-$(OH)_2D$. As vitamin D status does not affect DBP levels, the correlation suggests that 1,25-$(OH)_2D$ levels are related to changes in DBP concentration and that the free, rather than the total, 1,25-$(OH)_2D$ concentration is feedback regulated.

Despite a 1,000-fold difference between the total concentrations of the liver metabolite 25-hydroxyvitamin D and 1,25-(OH)D, the difference between the free concentrations and these metabolites is only 10-fold (Bouillon). If it is the free hormone which is important, it is clear that some thought will have to be given to the value of the current assays which measure total concentrations of vitamin D metabolites.

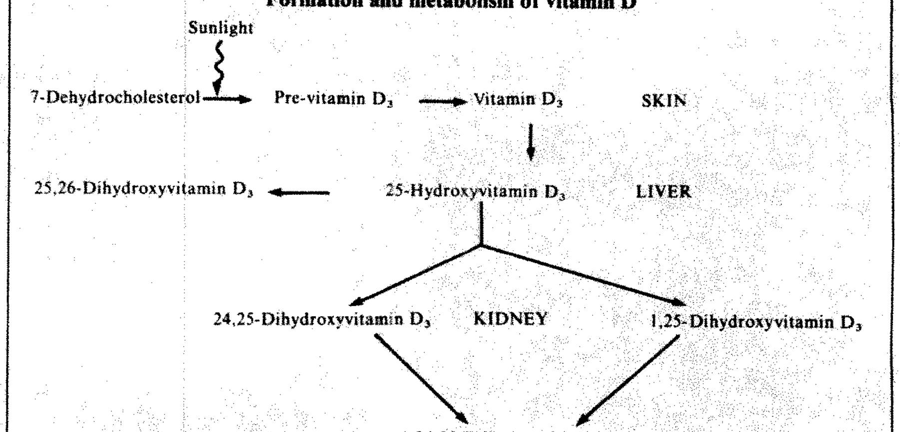
Another function of 1,25-$(OH)_2D$ which is becoming increasingly controversial is the hormone's action on the gut, where it increases the active transport of calcium. R.T. Franceschi and H.F. De Luca (*J. biol. Chem.* **256**, 3848; 1981) and D.E.M. Lawson (Dunn Nutrition Laboratory, Cambridge) believe that it is new proteins synthesized in the mucosa in response to 1,25-$(OH)_2D$ which facilitate calcium transport across the gut. But H. Rasmussen (Yale University) argues that the increase in calcium transport is not affected when protein synthesis is inhibited. He has observed that 1,25-$(OH)_2D$ increases the synthesis of phosphatidylcholine in mucosal cells — in the absence of protein synthesis — and he believes that this change is necessary for active calcium transport.

Rasmussen claims that the increased phosphatidylcholine concentration is eventually reflected in changes in the phospholipid composition of the brush border membrane. D.D. Bikle (University of California) would probably support this view. He ruled out newly synthesized protein within brush border membrane, or altered phosphorylation of proteins in response to 1,25-$(OH)_2D$, as the explanation for the increased calcium flux across the membrane. He found that calcium transport occurred in the absence of both these changes.

J.A. Putkey (University of California), however, could find no evidence to support the lipid hypothesis. Use of a spin label probe, 5-nitroxide stearate, showed that dietary vitamin D has no detectable effect on either the polarity, or lipid fluidity, of the membrane; neither does it alter its cholesterol or lipid phosphorus content. It seems, however, that some consensus view, which relies on both protein synthesis and lipid turnover to explain the action of 1,25-$(OH)_2D$ in the gut, is sure to emerge sooner or later.

If just one or two of vitamin D's metabolites can generate so much controversy, then the chemists would seem to have provided ample ammunition to allow the disputes to continue for some time yet. □

Formation and metabolism of vitamin D



*The 5th Workshop on Vitamin D was held at Historic Williamsburg, Virginia on 14–19 February 1982. Proceedings will be published by Walter de Gruyter later this year.

Alastair Hay is a Lecturer in the Department of Chemical Pathology, University of Leeds, Leeds LS2 9JT.

REVIEW ARTICLE

Variety in the level of gene control in eukaryotic cells

James E. Darnell Jr

The Rockefeller University, New York, New York 10021, USA

In light of present progress in the understanding of how messenger RNA is constructed in eukaryotic cells, the levels of gene control are discussed. Transcriptional control, assessed by the rate of synthesis of specific nuclear RNA, is strongly indicated to be the most frequent level of control. Definition of eukaryotic transcription units as simple (encoding one protein) and complex (encoding two or more proteins) affords a framework in which to discuss control at the level of RNA processing for which several examples also are known. Finally, differential stability of cytoplasmic mRNAs and differential translation, both well established, are briefly described.

THE study of gene control in eukaryotic cells has matured over the past five years into one of the most active and productive fields in experimental biology Earlier conclusions about eukaryotic gene control were deferred because mRNA production in eukaryotes is not a simple matter of transcribing the RNA Especially in vertebrate cells, the transcription units for mRNAs are longer than the final mRNA product[1-6] Several 'processing' steps occur to primary RNA transcripts in the cell nucleus before the mRNA emerges to be translated in the cytoplasm, involving the addition of a methylated guanylate residue—a 'cap'—at the 5' end of primary transcripts[7], the addition of adenylic acid residues to a site in the primary transcript to create a 3' poly(A) segment[8-13] and finally the removal of specified internal sequences (introns)[14] and rejoining or splicing of the remaining RNA pieces (exons)[14] to create a finished mRNA (refs 15-17, Fig 1)

This complicated mRNA formation, together with the ordered, complex structure of the nucleus and cytoplasm of eukaryotic cells and the different fundamental needs served by a genetic programme in slowly growing or non-growing differentiated eukaryotic cells and rapidly adaptable bacterial cells, have long spurred speculation[5,18,19] that eukaryotic gene control might differ from that in bacteria In particular, it has been suggested that eukaryotic genes are subject to regulation at more levels than are bacterial genes In addition to transcriptional regulation, there might be regulation of RNA processing in the nucleus, mRNA transport from the nucleus, and mRNA stability and frequency of translation in the cytoplasm

With the major mechanical steps of eukaryotic mRNA formation now understood, it is possible to describe more accurately the anticipated levels and varieties of eukaryotic gene control[5] In fact, regulation of the amount of specific mRNA molecules has been demonstrated at the level of transcription, nuclear RNA processing and mRNA stability Here I shall cite evidence supporting a variety of levels of eukaryotic gene control The experiments reviewed mainly demonstrate the levels of control mRNA concentration, and not the biochemical mechanisms underlying the controls No attempt is made to review the control of transcription units encoding ribosomal RNA, transfer RNA or 5S RNA (RNA polymerase I and III transcription units, transcription of 5S RNA has been recently reviewed[20]) The relationship between gene control and chromatin structure is not discussed but has been the subject of a review recently published[21] It is clearly possible to envisage gene control at the level of differential translation of mRNAs and potential differential post-translational events, but these subjects have received less experimental attention, so they will be mentioned only briefly

RNA transcription and transcriptional control

The first level of control in eukaryotic cells is in the choice of RNA polymerases In contrast to bacteria, eukaryotes have three RNA polymerases polymerase II transcribes genes that encode proteins[22,23], polymerases I and III have already been mentioned Both *in vivo*[24] and *in vitro*[23,25], RNA polymerase II initiates transcription at a site located at least in part by a conserved 8-10 nucleotide region 25-30 nucleotides upstream (in the 5' direction) from the RNA start site The nucleotides TATA are strongly conserved at this site[23,26] and the region is known as a 'TATA' or Goldberg-Hogness box[27] Undoubtedly other sequences are important in regulating access of polymerases to DNA Reports on viral, yeast and individual mammalian genes point to the importance of sequences upstream

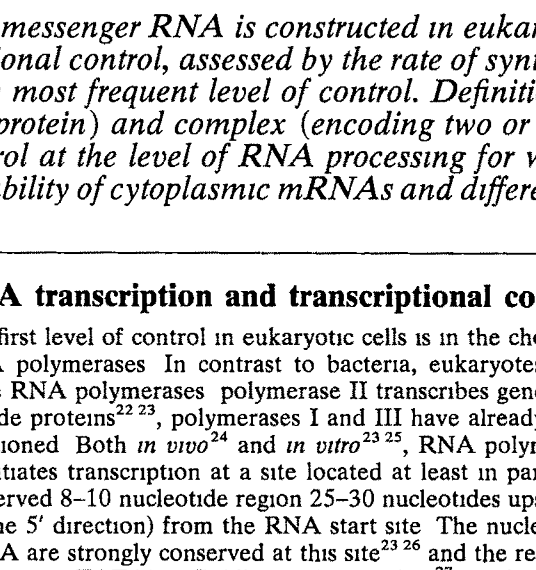


Fig. 1 Steps in mRNA formation (1) DNA indicating the sites of initiation (INIT), about 30 nucleotides to the 3' site of a TATA box, and of termination (Term) of transcription (2) Initiated transcript with cap (m^7Gppp) added to the first nucleotide (N) (3) Transcript after polyadenylation—the addition of a sequence of about 200 adenylate residues (A_{200}) at the poly(A) site, which is 10-20 nucleotide bases downstream from the conserved hexanucleotide AAUAAA Between steps (2) and (4), mammalian mRNAs acquire methyl groups ($-CH_3$) on the 6 position of about 1 of every 100 adenylate residues The exact time of methyl group addition is not known (4) Messenger RNA formed by splicing X, poly(A) site, s, splice site

0028-0836/82/220365—07$01 00

from the TATA box (see, for example refs 28–33) Although no consistent picture has yet emerged for the role of these upstream sequences in transcriptional control, intense research in this area should yield important conclusions soon

Here we are concerned more with establishing the variation of transcription rates than with the basis for the variation Transcription initiation is widely believed to be the major controlling step in transcriptional control Little is known, however, about transcriptional termination, which might also be a level of control For example, all RNA polymerase II complexes that begin a chain do not transcribe the entire chain Prematurely terminated transcripts are produced from viral[34] and probably cell[35] transcription units Little is known about the specificity of transcriptional termination sites although polymerases seem regularly to pass poly(A) sites[10–12] Recent evidence suggests that, for the major late adenovirus[10,11] and the major β-globin transcription units[1,50] there exists a region of termination, if not a specific site

Transcriptional control

The various mechanisms for transcriptional control are very familiar from the classic studies of bacterial gene regulation[36,37] In bacteria, both regulated increases and decreases of transcriptional initiation are possible and two types of regulation of termination can occur—'readthrough' to the next gene and premature termination or 'attenuation' which stops transcription before the coding portion of the gene is reached[38] Apparent control of transcriptional initiation that results in an increase or decrease of mRNA has been documented in several cases in eukaryotic cells and one case of differential termination has been described The evidence reviewed here shows clearly that the most frequent mode of eukaryotic gene control is at the transcriptional level

Measurement of transcription control. Bacterial geneticists used RNA–DNA hybridization to document the existence and very short half life of mRNA[39] and to map transcription units[37,38,40,41] The simple presence or absence of specific mRNAs was taken as evidence (correctly in most cases) for transcriptional control However, in eukaryotic cells, a changing level in the steady-state concentration of a given mRNA or an increased labelling of mRNA after relatively long label times could partly be due to differential nuclear or cytoplasmic stabilization To avoid this difficulty it is necessary to hybridize very briefly labelled ('nascent') nuclear RNA to DNA sequences corresponding to a specific mRNA[5] The ideal method to label nuclear RNA is expose cells to radioactive RNA precursors for say ≤ 5 min, and thereby sample the labelled nuclear RNA even before the acid-soluble triphosphate pool has become fully labelled[10,35] But this is often not practicable because so little of the total RNA output (usually 5×10^{-6}–5×10^{-5} of the total) from the nucleus is due to transcription of a single gene[4,42,43] Fortunately there is another suitable method for labelling nascent nuclear RNA In isolated nuclei, previously initiated RNA polymerase II molecules elongate growing RNA chains by less than 500 nucletides in 10–20 min of incubation[1,4,44] RNA synthesis in isolated nuclei depends on nucleoside triphosphates (NTPs) and thus nuclear RNA can be highly labelled with high-specific activity NTPs Two studies suggest that *in vitro* chain elongation is an accurate reflection of a brief label inside cells First, in the cases of adenovirus-2 transcription units[1] and the major β-globin transcription unit[4], the polymerases in isolated nuclei elongate intact molecules covering the entire transcription unit Second, the fraction of total *in vitro* or *in vivo* labelled RNA complementary to various specific DNA segments is similar[1–4,45–48]

Therefore isolated nuclei are widely used to prepare nascent RNA for the measurement of differential transcription of specific genes All the experiments I cite as establishing transcriptional control have used RNA from either pulse-labelled cells or isolated nuclei to estimate transcriptional rate

Adenovirus transcriptional control. The changes in virus protein and virus mRNA synthesis during adenovirus infection have been thoroughly explored (Fig 2) Specific increases and decreases in rates of transcription, presumably mediated at the initiation step, occur at several sites on the adenovirus genome during virus infection In addition, one case of apparent termination control has been described The following conclusions can be drawn about adenovirus transcriptional controls

(1) During the first 1–2 h of infection, transcription units 1A, 1B, 3 and 4 all increase in transcription rate[49] Of this group, transcription of 1A appears to peak first In comparison, transcription unit 2 is delayed[49,50] and the transcription unit for protein IX produces no mRNA or nuclear RNA at all for the first 4–6 h (refs 51, 52), although the protein IX transcription unit resides entirely within the early transcription unit 1B (refs 26, 52, 53, Fig 2) At ~6–8 h after infection, protein IX transcription increases without affecting the rate of transcription of 1A or 1B (ref 52) Among the tests for early activity of the protein IX transcription unit was an assay for promoter-proximal RNA capable of detecting minimal (~1%) synthesis of the first 100 bases of the transcription unit, but still no activity was evident[52] This is one of the clearest cases in eukaryotic cells or viruses of transcriptional control at the level of initiation

(2) Mutations in transcription unit 1A lead to a failure during infection of mRNA accumulation from regions 1B, 2, 3 and 4 (refs 54–57) This failure is due to the absence of transcription when a 1A protein product is defective[57], but the exact mechanism of action of the 1A protein remains unknown[52]

(3) P16 (the major late adenovirus promoter) functions both early[59–62] and late[1,3,24,34] in infection, but after DNA synthesis, it is 100–1,000 times more active per cell[34,60] This increased transcription is presumably on new DNA templates In comparison, transcription units 1A and 1B on the same strand remain equally active throughout infection[34,52,63] Thus there seems to be a mechanism for restricting polymerase initiation to P16 on the new DNA

(4) Controlled termination of transcription also occurs during adenovirus infection[60–63] The same P16 promoter produces RNA beginning at the same cap site both early and late in infection From the late transcripts, five 3' co-terminal groups of mRNA are formed[10,64–66], but early in infection only the L1 poly(A) site is found in mRNA[60,62] Early transcription does not seem to pass map unit 60 and perhaps ends between 40 and 50 (refs 60–62), whereas late transcription continues to around map unit 98 (refs 10, 11) This is equivalent to readthrough transcriptional control[37,38] As there is an enormous increase in transcription from P16 when DNA is replicated it seems likely that transcription of the same DNA molecule is

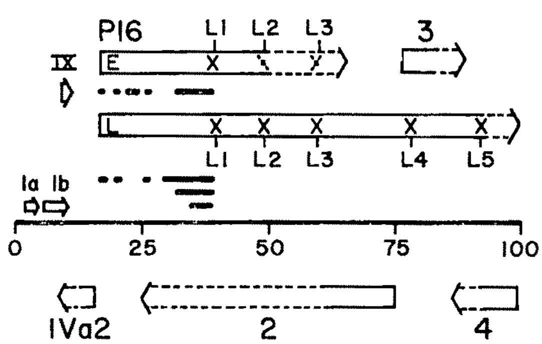


Fig 2 Features of the transcription map of adenovirus transcription units (adapted from refs 5, 6, 52, 59–63) The genome is 36 kb long, so each map unit equals 360 bases Each open bar, 1a, 1b, 1X, 3, represents a transcription unit The transcription unit P16 beginning at map position 16 is shown both early (E) and late (L) in infection The arrowheads show direction of transcription and dashed lines indicate cases where transcription proceeds beyond poly(A) sites Only for P16 are the poly(A) sites marked(X) Special prominence is given to P16 to show that early in infection, transcription definitely terminates before reaching L3, L4 and L5, the thick black lines under P16E and P16L indicate the different regions retained after splicing in early and late L1 mRNAs

not terminated in two different ways, but that one transcription unit in the adenovirus genome can have different termination sites early and late in infection

(5) Transcription of region 4 peaks early in infection and then at 4–8 h declines to ≤10% of the rate it reached by 4 h (ref 67) Protein synthesis is necessary for this programmed decline, particularly a protein from region 2 as mutations in region 2 block the decrease

SV40 transcriptional control. A specific decrease in the early transcription unit of simian virus 40 (SV40) has been documented during the later periods of infection[68] At the beginning of infection, the overall rate of transcription of SV40 is quite low and there is five times as much RNA complementary to the early transcription unit as to the late transcription unit Late in infection, after a very large increase in total SV40 RNA in infected cells, there is only 1/20th as much RNA complementary to the early transcription unit as to the late transcription unit This change in ratio of mRNAs from the two transcription units is probably due to the T antigen, the protein encoded by the early transcription unit[69] T antigen has been shown to bind three identical sites within a 100–150 base pair region that overlaps the early promoter[70,71] This binding is necessary for DNA replication[72] but also stops or slows considerably the early transcription of SV40 The effect of T antigen is specific to the early promoter as transcription of recombinant DNA molecules, in which the adenovirus major late promoter has been inserted no more than 50 bases away from the early SV40 promoter, is unaffected by the *in vitro* binding of the T antigen to the SV40 promoter[71]

Hormone-induced increases in transcription. Many proteins in differentiated cells are known to be produced in increased amounts in tissues stimulated by the appropriate hormone Transcription in isolated nuclei shows that this stimulation is due, at least mainly, to transcriptional control

One example is the stimulation of production of egg white proteins in the chicken oviduct by oestrogen cDNA clones to the egg white proteins ovalbumin[73,74], conalbumin[74] and ovomucoid[75] all hybridize 10–50-fold more labelled nuclear RNA from the nuclei of hormone-treated chick oviduct epithelial cells than from control cells Nuclei from other tissues such as liver or spleen do not make labelled RNA complementary to ovalbumin or ovomucoid cDNA Maintenance of a high level of transcription in the oviduct requires continuous hormone treatment

Liver cells of roosters also respond to oestrogen treatment by producing vitellogenin and very low density lipoprotein (VLDL)[76,77] The cDNAs corresponding to these mRNAs detect a greatly increased rate (>50-fold) of RNA synthesis in chromatin fraction (lysed nuclei) from treated compared with untreated animals Thus increased transcription of specific genes by steroid hormones seems well established

In addition to these cases of cell genes, RNA synthesis from the integrated mammary tumour virus DNA also is increased by oestrogens[29]

Transcriptional control of chick globin. For the first few days after fertilization, chick embryos synthesize embryonic forms of α- and β-globin Later on adult forms of the two globins are synthesized The same cell does not differentiate first to form embryonic and then an adult globin Rather, erythroblastic cells choose one of the two pathways[78,79] Genomic DNA clones have been used to show that isolated nuclei from 5- or 12-day-old embryos produced labelled RNA complementary only to embryonic or adult globin DNA respectively Thus the differential expression of these two forms of chick globin appears to be controlled at the level of transcription

Liver-specific transcriptional control. A series of cDNA clones prepared from mouse liver mRNA were screened to select mRNA sequences present in many cell types ('common' mRNAs) and mRNA sequences present only or mainly in liver[46] Nuclei from liver cells, cultured undifferentiated cells and brain cells were then used to prepare labelled nascent RNA All 11 liver-specific mRNA sequences were labelled in liver nuclei but not in brain or cultured cell nuclei Sequences complementary to the common cDNA clones were labelled in all nuclei As the clones used in the test represented mRNAs that ranged from ~100 to >10,000 copies per cell, it seems that, in general, specific mRNAs in liver cells are controlled transcriptionally A further experiment with liver nuclei reveals a specific case of changing transcriptional controls Adult male mice excrete a protein in urine, the major urinary protein (MUP), which is not excreted by newborn mice[80] Liver nuclei from adult and newborn animals were used to score transcription of MUP genes The adult nuclei synthesized at least 100-fold more MUP-specific RNA than did the nuclei from the livers of newborns[81]

Transcriptional control by choice of initiation sites. More of the same α-amylase enzyme protein is produced in rat salivary glands than in liver When the cDNAs complementary to α-amylase mRNA from the two tissues were sequenced, the 5′ nucleotides of the two mRNAs were found to differ while the remainder of the mRNA chains were identical[82] Examination of genomic DNA[83] and nulcear RNA (U Schibler, personal communication) revealed that the cap sites for the two amylase mRNAs lie ~2 8 kilobases apart but the two overlapping transcription units cover the same protein coding region and the same poly(A) site(s) This necessitates a different splice to join the two different cap and leader sequences to the same mRNA 'body' However, as cap sites always seem to be start sites for RNA synthesis[4,23–26], α-amylase production would appear to be controlled by selecting specific transcriptional start sites Different overlapping transcription units in adenovirus (1b and protein IX, and region 3 and P16) are also regulated by choosing the proper start sites (ref 41, Fig 3), as are the induced and constitutive forms of yeast invertase[151]

Transcriptional control dependent on DNA rearrangement. Such interest has attended the possibility that rearrangements affect gene control that the proved examples of such rearrangements must at least be mentioned Immunoglobulin formation, the variation in surface glycoproteins of trypanosomes, switching of the yeast mating-type locus and promoter insertion brought about by retrovirus movement in chromosomes all appear to be cases where gene activation, probably transcriptional activation, depends on DNA rearrangement Actual transcriptional rate studies have not been done in most of these cases (see ref 84)

RNA processing and its control

Although transcriptional control may be the predominant level of gene control, it is not the only one To help understand post-transcriptional control, it is best first to consider more fully the steps in mRNA manufacture (see Fig 1)

The sequence of events that convert a primary RNA transcript to a mRNA are collectively known as 'RNA processing' These events begin very soon after initiation of an RNA chain Before the new RNA chain is longer than 50 nucleotides[34,85], and perhaps before the RNA polymerase II moves past the initiation site, a 'cap', a m^7Gp residue[7], is added to the first nucleotide (N) The middle phosphate of the m^7GpppN cap structure is not derived from the GTP but from the pppN that starts the RNA chain[86–88], strong evidence of RNA chain initiation at the cap site Few, if any, polymerase II products escape this 5′ end addition, even prematurely terminated RNA chains[33,34,85] and chains synthesized *in vitro* by cell extracts supplemented with RNA polymerase II[25] are capped

The poly(A) segment is added post-transcriptionally to heterogeneous nuclear RNA (hnRNA) in the cell nucleaus[8,9,20] Although the precision of termination sites is not yet understood for polymerase II transcription, the enzyme has been shown to transcribe past the site for poly(A) addition in at least three adenovirus transcription units[1,10–12], both early and late SV40 (refs 13, 94) and polyoma transcription units[95,96], the β-globin transcription unit in mouse cells[4] and chicken cells[78,79], and probably the transcription unit for the heavy chains of immunoglobulins[97–99] No clear cases of termination at the poly(A) site

have been documented by studying briefly labelled nuclear RNA Thus addition of poly(A) may always require endonucleolytic cleavage and terminal addition of poly(A) to the cleaved 3′ end of a nuclear RNA molecule At least part of the signal for poly(A) addition is the AAUAAA sequence found 10–25 bases upstream from the poly(A) addition site[100 101]

Poly(A) addition usually, if not always, precedes in time the final, splicing[15–17] steps of mRNA processing[10 102–104] Splicing, however, can occur in the absence of poly(A) addition[105] and mRNA formed during inhibition of poly(A) synthesis can at least in part enter polyribosomes[106], although it does not accumulate there as stable mRNA These recent results suggest that the major (or perhaps sole) role of the poly(A) is cytoplasmic stabilization of the mRNA[106–108] and that splicing for most transcripts might be an automatic event Methylation of mRNA[7], which is known to occur (on the average) to 1 out of 400 Ap residues in mammalian mRNAs[109–111], may occur before splicing, but its timing and role in mRNA formation remain speculative[111] Finally mRNAs leave the nucleus and enter the cytoplasm still bearing their poly(A) segment[8 19] With time, the poly(A) tail shortens[8], which causes some mRNA sequences, >90% of which initially partition experimentally in the poly(A)-containing fraction, to fractionate in the poly(A)-lacking fraction[92 93] Extensive attempts with sea urchin material have failed to isolate any cloned DNA segment complementary to the mRNA molecules of the latter fraction that are not also present in the poly(A)-containing fraction (M Nemer, personal communication) other than histone mRNA[89]

Transcription unit design and differential processing. Two categories of transcription units can be identified (Fig 3) (1) Simple transcription units encode a single protein, whether or not processing of the primary transcript is required Histone genes are simple transcription units that do not, in most cases, require either poly(A) addition or splicing[90] Protein IX mRNA of adenovirus[112] and the α-interferon mRNAs[113] derive from simple transcriptional units whose products have poly(A) added but do not require splicing The α- and β-globin transcription units are simple because, although their primary transcripts require poly(A) addition and splicing, only one functional mRNA is formed from the transcription unit[114 115]

(2) Complex transcription units are those whose primary transcript can give rise to two or more mRNAs that encode two different proteins Two different structural arrangements are known for these units two or more poly(A) sites and two or more variations in splicing patterns Both virus and cell transcription units with multiple poly(A) sites are known—the major late adenovirus transcription unit[10 63–65], the heavy chains of immunoglobulins[97–99] and the transcription unit encoding the peptide hormone calcitonin (R Evans and G Rosenfeld, personal communication) Thus far only virus transcription units are known that have one poly(A) site and two or more splicing sites—both transcription units of SV40 (refs 94, 117) and polyoma[95 118], most of the nine known adenovirus transcription units[16 58 116 119] and many different retroviruses[120] A number of virus transcription units with multiple splicing arrangements can be integrated into cell DNA and still produce multiple mRNAs Thus, it is clearly possible for DNA in the cell genome to encode primary transcripts that can be spliced in alternative ways (Transcription units can have two or more poly(A) sites and variable splicing at one or both poly(A) sites, such as in the major late adenovirus transcription unit, Fig 2)

From these two types of complex transcription unit follow two general types of processing control[5] (1) Differential processing each transcript from a complex transcription unit must be processed in one of two or more ways Several cases are known in which various poly(A) sites or splice sites can be chosen on the same primary transcript (2) Process versus discard decisions In this hypothetical possibility, there would be little or no processing of a particular transcript in one cell type or a cell in one particular physiological condition but successful processing of the same primary transcript in another cell type or another physiological condition (reviewed in ref 5) Candidate transcripts of this type do exist in cultured cells and probably also in tissues but no specific case of gene control through process versus discard decisions is yet proved

Adenovirus processing control. A differential choice of both poly(A) and splicing sites has been documented in the primary transcript of adenovirus mRNA, P16 (major late promoter, Fig 2) Differential choice of poly(A) sites is also suspected to occur in at least one other transcription unit (region 2 transcripts that lead to mRNA formation with sequences in the 20–30 map unit region)[58 121]

Early in infection, in transcripts beginning at P16, the poly(A) sites at both 38 5 (L1) and 50 (L2) map units are used to generate poly(A)-containing nuclear RNA[59–61] About two L1 molecules are produced for each L2 Late in infection, in addition to the whole transcription unit being transcribed and the L1–5 being available for use, a change occurs so that L2 is used about twice as often as L1 (ref 10) This change in ratio of poly(A) site usage accords with the change in amount of L1 versus L2 mRNAs in the cell nucleus Late in infection, when five poly(A) sites are available, L3 is more frequently chosen than any other site but the ratio of L1–5 remains at 1 2 3 2 2 throughout the late course of infection (ref 10 and J R Nevins, unpublished)

In addition to differential poly(A) choices, the early splicing pattern is different from the late one in the L1 poly(A)$^+$ nuclear molecules Examination of nuclear RNA indicates clearly that three products are made late in infection but only one major product early in infection[60] In addition, the initially labelled cytoplasmic molecules late in infection contain all three species of L1 mRNA whereas early in infection only a single species of labelled mRNA emerges in the cytoplasm after a brief label[60] Finally, the protein translated from the early L1 mRNA is encoded mainly, if not exclusively, in the 400 nucleotides that are retained in early L1 mRNA but are lost in the late mRNA[61] Thus the evidence seems very strong that a change in splicing takes place in the handling of the same sequence of nucleotides contained in the nuclear poly(A)$^+$ precursor molecule that is encoded in the 16–38 5 region on the adenovirus genome

Immunoglobulin mRNA formation. Only two cases of differential processing are known for cellular mRNAs During B-lymphocyte development and immunoglobulin synthesis, the μ heavy chain is first inserted as an integral protein in the plasma membrane[97–99] Later a similar μ chain is found as part of secreted immunoglobulins Analysis of cDNA clones of the two

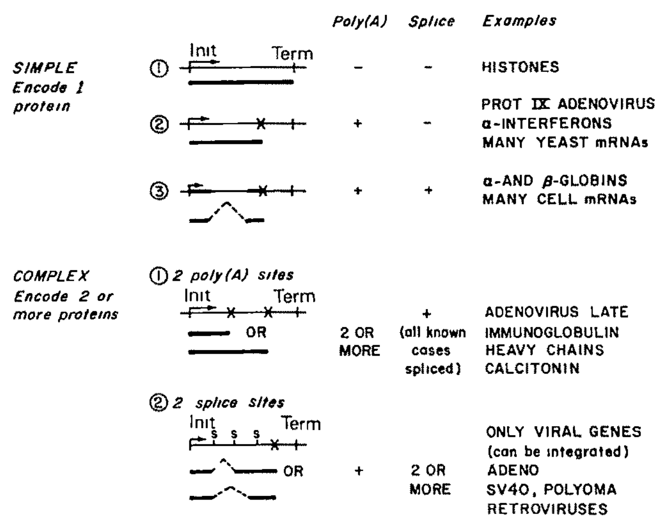


Fig. 3 Definition and examples of simple and complex transcription units In each case, DNA is shown at the top and the mRNA(s) underneath On the former are indicated sites of initiation (INIT) and termination (TERM) of transcription, splicing sites (s) and the position of the sequence which will become the site of poly(A) addition to the processed transcript

mRNAs showed that the deduced carboxyl terminus of the μ membrane and μ secreted chains was different but the deduced amino termini were the same Analysis of the genomic cloned DNA revealed two poly(A) sites in the same transcription unit and the primary transcript seems to cover both of them (ref 97 and R Wall, personal communication) Depending on the choice of poly(A) site two μ heavy chain mRNAs with different 5′ most sequences can be formed, one for the μ membrane chain and one for the μ secreted chain Further studies suggest that the same five exonic regions common to both μ chains also appear in δ chains[99] In this case, the 3′ exonic component of the μ mRNA may be a consequence of choosing a poly(A) site more than 25 kb downstream These studies were carried out on lymphocytic tumour cells in culture, but it is thought that a similar set of events occurs in the normal differentiation sequence in B cells

Note that the molecular basis for none of these presumed differential choices of poly(A) and splice sites is known and it remains possible that template changes may dictate the post-transcriptional outcome For example, rather than there being a single completed transcript in which a poly(A) site is chosen, a selective protein that bound to the DNA at one or the other poly(A) site might instead trigger the choice of poly(A) site during transcription

Process versus discard decisions. Differential processing of nuclear RNA from either simple or complex transcription units might include the processing or discard of a particular nuclear RNA sequence[5 18–20] This type of processing is different from differential poly(A) choices or differential splicing choices where a transcript has a high probability of yielding one or another mRNA[10] Although there is no evidence of variable process versus discard between two different cells, nuclear RNA turnover is substantial[122] and nuclear RNA 'complexity' (average nucleotide content) is some 10–20 times greater than the complexity of the mRNA[123], perhaps greater than the difference expected on the basis of intron loss during nuclear RNA processing

More recent analyses of nuclear RNA in cultured cells divide the nuclear RNA primary transcripts into two classes—those that produce mRNA sequences and those that do not While all transcripts contain a cap[34] and >90% of newly capped mRNAs that enter polyribosomes (excluding histone mRNAs) contain poly(A)[42 91], poly(A) is added in the nucleus only to about one-quarter of the total primary transcripts[91] Thus the rate of synthesis of caps is about four times that of poly(A) Moreover, when sequences complementary to eight cloned CHO cDNAs were examined in the poly(A)$^+$ and poly(A)$^-$ nuclear RNA fractions[24 42], whereas 70% of the sequences complementary to these clones were in the poly(A)$^+$ fraction, only 20–25% of the capped molecules were in the poly(A)$^+$ fraction It therefore seems that capped poly(A)$^-$ molecules whose sequences are not part of stable mRNAs make up a significant part of the nuclear RNA in cultured cells Can these primary transcripts at some other time and place (embryonic cells? specifically differentiated cells?) be shifted from the poly(A)$^-$ to the poly(A)$^+$ category and then appear as stable mRNA[5 20 125]? Specific clones complementary to these 'nucleus-confined' RNAs molecules are necessary to answer this possibility

Cytoplasmic gene control

Too little is known about nuclear RNA transport to suggest the existence of differential transport After the mRNA has reached the cytoplasm, however, the fates of mRNA molecules differ Evidence exists for both a wide range of half lives for different specific mRNAs in the same cells[62 92] and different half lives for the same mRNA in the same cell under different circumstances

Of these possible control mechanisms the first would, of course, allow the cell to accumulate differential concentrations of two different mRNAs although the synthesis rates of the two were identical, but its physiological significance remains obscure Changes in the half life of specific mRNAs under different circumstances has attracted more attention Two cases involving hormones are well documented

Breast tissue can be explanted and continue to respond to hormonal stimulation by the production of casein mRNA and casein protein[126] When explants are cultured without prolactin for 48 h they lose 95% of their casein mRNA but can, when given prolactin, accumulate about 25,000 copies of the casein mRNA in 24 h This accumulation is associated with an ~20-fold increase in the mRNA half life, but only a two- to threefold increase in the synthesis of casein mRNA sequences assayed by hybridization to casein cDNA

A second reported instance of mRNA stabilization occurs with oestrogen treatment of roosters[76,77] As described earlier, mRNA sequences, for both VLDL and vitellogenin are synthesized at least 20–100-fold faster in liver nuclei from treated than from untreated animals The rate of mRNA accumulation is maximal within about 6–8 h and from the curve of accumulation of vitellogenin and VLDL mRNAs during the first week of oestrogen treatment a half life of >24 h can be calculated If, after the mRNA has reached a plateau value, the oestrogen is abruptly withdrawn, the vitellogenin and VLDL mRNA decline with a half life of ~3 h Thus the half life changes from 24 to 3 h by removal of the hormone Calculations of similar hormone-dependent changes in half life for steriod-induced hormones have also been reported[127], although the changing steroid levels were not so carefully controlled

Differential half life of specific mRNAs also occurs during the growth cycle of adenovirus[62] Transcriptional unit 1B produces two mRNAs—one 20S and one 14S Early in infection the 20S form predominates while late in infection the 14S is predominant[129] The transcription rate of 1B sequences is constant from 5 to 24 h of infection[62] When labelled nuclear RNA was examined early and late in infection no changes were noted, suggesting that differential splicing was not occurring[62] However, cytoplasmic accumulations of 14S relative to 20S was greater late in infection than early in infection, pointing to an increased half life of the 14S mRNA This mRNA appears to have the same cap site, the same splicing and poly(A) sites late in infection as it does early and thus the same mRNA has a different half life at two different times in the infectious cycle

Finally, there are two established cases (and many others suspected) where preservation of a specific group of mRNAs occurs in the face of destruction of a large fraction of total cell mRNA[130–132] During erythroblast differentiation in mammals, the developing cell makes many different mRNAs including, in the final stages, globin mRNA Globin RNA synthesis,

Table 1 Summary of frequency of various types of control

Nuclear	Examples proved or strongly suggested	Possible
Transcriptional		
Initiation	Many, >100	
Termination		
Premature (attenuation) readthrough	1	+
Nuclear RNA processing		
Poly(A) choice	~3	
Splicing choice	1	
Process versus discard		+
Cytoplasmic		
mRNA stability	~5 Specific, many general	
mRNA translation efficiency	~10 Specific, many general	

For those cases of control where only one or a few cases are proved, it is anticipated that these numbers will increase It is not possible at present to even guess at the true frequency with which various mechanisms will be found ultimately to be used

however, represents a tiny fraction of total nuclear RNA synthesis During the final four or so cell divisions the cell becomes a highly specialized reticulocyte synthesizing over 90% globin plus a limited number of other specific proteins These differentiative steps appear to represent conservation of the globin (and a few other erythrocytic) mRNAs and specific destruction of all other mRNAs[130–132] A second case of specific mRNA destruction has been described during development of the slime mould *Dictyostelium discoideum* Cells which have aggregated will, if left alone, form a stalk, fruiting bodies and spores This is accomplished with the aid of a set of new RNAs formed after aggregation If the developmental cycle is interrupted after aggregation, the new developmentally regulated mRNAs are specifically destroyed[133]

Control of translation

Differential translation of finished mRNA molecules that are apparently already in the cell cytoplasm is recognized early in embryonic development, both for the average rate of protein synthesis per ribosome and for specific protein species Unfertilized sea urchin eggs become at least 50-fold more active in protein synthesis from the same number of mRNAs after fertilization than before[4] By contrast, in clam eggs the total rate of protein synthesis is unchanged after fertilization, however two-dimensional gel analysis of proteins newly formed *in vitro* compared with translation *in vitro* reveals certain proteins are formed specifically only pre- or post-fertilization[135] although the mRNAs are present

Major shifts in total protein synthesis have also been recognized in non-embryonic cells Cultured mammalian cells[136,137] and insect cells[138] decrease total protein synthesis after a heat shock while the translation of certain 'heat-shock' proteins is probably enhanced Cultured mammalian cells undergo a three-fold decrease in protein synthesis during each mitosis[139], and in virus-infected cells the synthesis of host proteins can be interrupted in several different ways although the mRNA is still present[140–142] It is also possible[143], though unproved, that tissue-specific favouritism for the translation of specific mRNAs or classes of mRNA occurs

Control after translation

Several types of post-translational control can be envisioned that would affect the final concentration of particular proteins in given cells For example, as polypeptides exist that can be cut by proteases into different, overlapping polypeptides with different activities, control of proteolysis represents a possible level of gene control This apparently occurs in the posterior pituitary where cells can produce either adrenocorticotropin or β-lipotropin presumably from the same primary translation product[144,145]

Different fates for the same polypeptide, such as differential protein turnover or differential prosthetic group additions, could obviously alter the enzymatic profile of a particular cell Whether such differential post-translational events occur or are common remains to be explored

Conclusion

The evidence summarized in this article demonstrates that a variety of mechanisms determine the abundance of specific mRNAs in eukaryotic cells, with the major level of control usually being transcription in the cell nucleus Although the mechanism of the specific increases or decreases in transcription is not the subject of this review, it is widely believed that site-specific binding proteins either change chromatin structure or directly facilitate or prevent RNA polymerase attachment at specific sites on the DNA[21] Only a few cases of supposed transcriptionally active proteins are known in vertebrate cells or virus-infected cells but quite a number of regulatory loci in lower eukaryotes, for example yeast, are now identified genetically (see, for example, refs 146–148) As yeasts have mRNAs with short half lives[149], most regulation of their rates of protein synthesis is probably due to controlled rates of transcription Progress in the isolation of specific, transcriptionally active proteins in yeasts can be expected soon

In addition to transcriptional initiation, a change in termination has apparently been identified in one adenovirus transcription unit It is unknown whether differential termination will be a common mechanism for regulation of transcription, although evidence of frequent premature terminations[33,34] is tantalizing in this regard Note that transcriptional control for a particular gene does not preclude other levels of control

The second level of nuclear gene control is processing It is now clear that differential poly(A) choices and differential splicing can occur, but if a transcription unit has only one poly(A) site and only one possible spliced mRNA, then differential processing of the primary transcripts cannot occur Of great importance is the possibility of process versus discard control (Cloning of poly(A)$^-$ primary transcripts from one cell source will be necessary to determine whether such transcripts ever become poly(A)$^+$ transcripts in other cells)

Differential mRNA turnover seems fairly common as even at this early stage in the study of specific cell mRNAs, several clear-cut cases of differential turnover are known

Although multiple levels of gene control seem to exist, we are almost completely ignorant of the logic of gene control systems in eukaryotic cells or even whether most genes are ever subject to specific control Regulation at transcription would appear for many genes to be quite strict, regulation of mRNA turnover, in contrast, is probably much less strict What might distinguish genes that are tightly regulated from those that are not remains unknown Finally the mechanisms and molecules that execute the various types of control remain unknown Perhaps only by examining these regulatory molecules will the molecular logic of eukaryotic gene regulation become clear

1 Weber, J Jelinek, W & Darnell, J E *Cell* **10,** 611–616 (1977)
2 Goldberg, S , Weber, J & Darnell, J E *Cell* **10,** 617–621 (1977)
2a Goldberg, S , Schwartz, H & Darnell, J E *Proc natn Acad Sci U S A* **74,** 4520–4523 (1977)
3 Evans, R M *et al* *Cell* **12,** 733–739 (1977)
4 Hofer, E & Darnell, J E *Cell* **23,** 585–593 (1981)
5 Darnell, J E *Prog Nucleic Acid Res molec Biol* **22,** 327–353 (1979)
6 Ziff, E B *Nature* **287,** 491–499 (1980)
7 Shatkin, A J *Cell* **9,** 645–653 (1976)
8 Brawerman, G *CRC Rev Biochem* , 1–38 (1981)
9 Jelinek, W *et al* *J molec Biol* **75,** 515–532 (1973)
10 Nevins, J R & Darnell, J E *Cell* **15,** 1477–1493 (1978)
11 Fraser, N W , Nevins, J R , Z ff, E & Darnell, J E *J molec Biol* **129,** 643–656 (1979)
12 Nevins, J R , Blanchard, J -M & Darnell, J F *J moiec Biol* **144,** 377–386 (1980)
13 Ford, J P & Hsu, M T *J Virol* **28,** 795–801 (1978)
14 Gilbert, W *Nature* **271,** 501 (1978)
15 Berget, S M , Moore, C & Sharp, P A *Proc. natn Acad Sci U S A* **74,** 3171–3175 (1977)
16 Chow, L T , Gelinas, R E , Broker, T R & Roberts, R J *Cell* **12,** 1–8 (1977)
17 Klessig, D F *Cell* **12,** 9–23 (1977)
18 Scherrer, K & Marcaud, L *J cell Physiol* **72,** Suppl 1, 181–212 (1968)
19 Darnell, J E , Jelinek, W & Molloy, G *Science* **181,** 1215–1222 (1973)
20 Korn, L J & Gurdon, J B *Nature* **289,** 461–465 (1981)
21 Weisbrod, S *Nature* **297,** 289–295 (1982)
22 Roeder, R in *RNA Polymerases* (eds Losick, R & Chamberlin, M) 285–329 (Cold Spring Harbor Laboratory New York, 1976)
23 Corden, J *et al* *Science* **209,** 1406–1414 (1980)
24 Ziff, E & Evans, R M *Cell* **15,** 1463–1475 (1978)
25 Weil, P A , Luse, D S , Segall, J & Roeder, R G *Cell* **18,** 469–484 (1979)
26 Baker, C & Ziff, E *J molec Biol* **149,** 189–221 (1981)
27 Goldberg, M thesis, Stanford Univ (1978)
28 Benoist, C & Chambon, P *Nature* **290,** 304–310 (1981)
29 Ucker, D S , Ross, S R & Yamamoto, K R *Cell* **27,** 257–266 (1981)
30 Struhl, L & Davis, R *J molec Biol* **52,** 553–568 (1970)
31 Struhl K *J mo'ec Biol* **52,** 569–576 (1970)
32 McKnight, S L Gavis, E R , Kingsbury, R & Axel, R *Cell* **25,** 385–398 (1981)
33 Mellon, P , Parker, V , Gluzman, Y & Maniatis, T *Cell* **27,** 279–288 (1981)
34 Fraser, N W , Sehgal, P B & Darnell, J F *Proc natn Acad Sci U S A* **76,** 2571–2575 (1979)
35 Salditt-Georgieff, M , Harpold, M , Chen-Kiang, S & Darnell, J E *Cell* **19,** 69–78 (1980)
36 Jacob, F & Monod, J *J molec Biol* **3,** 318–356 (1961)
37 Miller, J H & Reznikoff, W W (eds) *The Operon* (Cold Spring Harbor Laboratory, New York, 1978)
38 Yanofsky, C *Nature* **289,** 751–759 (1981)
39 Hayashi, M , Spiegelman, S , Franklin, N C & Lauria, S E *Proc natn Acad Sci U S.A* **49,** 729–736 (1963)
40 Bertrand, K , Squires, C & Yanofsky, C *J molec Biol* **103,** 319–337 (1976)
41 Lozeron, H A , Dahlberg, J E & Szybalski, W *Virology* **71,** 262–277 (1976)
42 Harpold, M M , Evans, R M , Salditt-Georgieff, M & Darnell, J E *Cell* **17,** 1025–1035 (1979)
43 Salditt-Georgieff, M & Darnell, J E *Molec cell Biol* **3** (in the press)
44 Cox R F *Cell* **7,** 455–465 (1976)

45 Wall, R, Philipson, L & Darnell, J E Virology 50, 27–34 (1972)
46 Weber, J thesis, Rockefeller Univ (1979)
47 Derman, E et al Cell 23, 731–739 (1981)
48 Lowenhaupt, K, Trent, C & Lingrel, J B Devl Biol 63, 441–454 (1978)
49 Nevins, J R, Ginsberg, H W, Blanchard, J M, Wilson, M C & Darnell, J E J Virol 32, 727–733 (1979)
50 Carter, T H & Blanton, R J Virol 25, 664–674 (1978)
51 Pettersson, U & Matthews, M B Cell 12, 741–750 (1977)
52 Wilson, M C, Fraser, N W & Darnell, J E Virology 94, 178–184 (1979)
53 Fraser, N W & Ziff, E J molec Biol 155 (in the press)
54 Jones, N & Shenk, T Proc natn Acad Sci USA 76, 3665–3669 (1979)
55 Harrison, T, Graham, F & Williams, J Virology 77, 319–329 (1977)
56 Berk, A J, Lee, F, Harrison, T, Williams, J & Sharp, P A Cell 17, 935–944 (1979)
57 Nevins, J R Cell 26, 213–220 (1981)
58 Shenk, T, Jones, N, Colby, W & Fowlkes, D Cold Spring Harb Symp quant Biol 44, 367–375 (1979)
59 Chow, L T Broker, T R & Lewis, J B J molec Biol 134, 265–303 (1979)
60 Shaw, A R & Ziff, E Cell 22, 905–916 (1980)
61 Nevins, J R & Wilson, M C Nature 290, 113–118 (1981)
62 Akusjarvi, G & Persson, H Nature 292, 420–426 (1981)
63 Wilson, M C & Darnell, J E J molec Biol 148, 231–251 (1981)
64 Ziff, E & Fraser, N W J Virol 25, 897–906 (1978)
65 McGrogan, M & Raskas, H J Proc natn Acad Sci USA 75, 625–629 (1978)
66 Nevins, J & Darnell, J E J Virol 25, 811–823 (1978)
67 Nevins, J R & Jensen-Winkler, J Proc natn Acad Sci USA 77, 1893–1897 (1980)
68 Acheson, N H in Molecular Biology of Tumor Viruses 2nd edn, Pt 2 (ed Tooze, J) (Cold Spring Harbor Laboratory, New York, 1980)
69 Alwine, J C, Kemp, D J & Stark, G R Proc natn Acad Sci USA 74, 5350–5354 (1977)
70 Rio, D, Robbins, A, Meyers, R & Tjian, R Proc natn Acad Sci USA 77, 5706–5710 (1980)
71 Myers, R M, Rio, D C, Robbins, A K & Tjian, R Cell 25, 373–384 (1981)
72 Gutai, M W & Nathans, D J molec Biol 126, 259–274 (1978)
73 Swaneck, G E, Nordstrom, J L, Krenzalee, F, Tsai, M-J & O'Malley, B W Proc natn Acad Sci USA 76, 1049–1053 (1979)
74 McKnight, G S & Palmiter, R D J biol Chem 254, 9050–9058 (1978)
75 Tsai, S et al Biochemistry 17, 5733–5780 (1979)
76 Goldberger R in The Role of RNA in Development and Reproduction (eds Niu, M & Chuang, H H) 215–231 (Science, Beijing, 1981)
77 Wiskocil, R et al Proc natn Acad Sci USA 77, 4474–4478 (1980)
78 Groudine, M, Peretz, M & Weintraub, H Molec cell Biol 1, 281–288 (1981)
79 Groudine, M & Weintraub, H Cell 24, 333–334 (1981)
80 Finlayson, J S, Asofsky, R, Potter, M & Runner, C Science 149, 981–982 (1965)
81 Derman, E Proc natn Acad Sci USA 78, 5425–5429 (1981)
82 Hagenbuchle, O et al Nature 289, 643–646 (1981)
83 Young, R A, Hagenbuchle, O & Schibler, U Cell 23, 451–458 (1981)
84 Cold Spring Harb Symp quant Biol 46 (1981)
85 Babich, A, Nevins, J R & Darnell, J E Nature 287, 246–248 (1980)
86 Contreras, R & Fiers, W Nucleic Acids Res 9, 215–236 (1981)
87 Venkatesan S & Moss, B J J biol Chem 255, 2835–2842 (1980)
88 Venkatesan, S, Gershowitz, H & Moss, B J J biol Chem 255, 2829–2834 (1980)
89 Adesnik, M & Darnell, J E J molec Biol 67, 397–406 (1972)
90 Hentschel, C C & Birnstiel, M L Cell 25, 30–313 (1981)
91 Salditt-Georgieff, M, Harpold, M, Wilson, M & Darnell, J E Molec cell Biol 1, 179–187 (1981)
92 Harpold, M, Wilson, M & Darnell, J E Molec cell Biol 1, 188–198 (1981)
93 Minty, A J & Gros, F J molec Biol 139, 61–83 (1980)
94 Reddy, V B et al Science 200, 494–502 (1978)
95 Legon, S, Flavell, A, Cowie, A & Kamen, R Cell 16, 373–388 (1979)
96 Acheson, N H Proc natn Acad Sci USA 75, 4754–4758 (1978)
97 Rogers, J et al Cell 20, 303–312 (1980)
98 Alt, F W et al Cell 20, 293–301 (1980)
99 Maki R et al Cell 24, 353–365 (1981)
100 Proudfoot, N J & Brownlee, G G Nature 263, 211–214 (1976)
101 Fitzgerald, M & Shenk, T Cell 24, 251–260 (1981)
102 Lai C-J, Dhar, R & Khoury, G Cell 14, 971–982 (1978)
103 Tilghman, S M, Curtis, P J, Tiemeier, D C, Leder, P & Weissman, C Proc natn Acad Sci USA 75, 1309–1313 (1978)
104 Roop, D R et al Cell 15, 671–685 (1978)
105 Zeevi, M, Nevins, J R & Darnell, J E Cell 26, 39–46 (1981)
106 Zeevi, M, Nevins, J R & Darnell, J E Molec cell Biol 3 (in the press)
107 Nudel, U et al Eur J Biochem 64, 115–125 (1976)
108 Huez, G, Bruck, C & Cleuter, Y Proc natn Acad Sci USA 78, 908–911 (1981)
109 Salditt-Georgieff, M et al Cell 7, 227–237 (1976)
110 Sommer, S et al Nucleic Acids Res 3, 749–766 (1976)
111 Chen-Kiang, S, Nevins, J R & Darnell, J E J molec Biol 135, 733–752 (1979)
112 Alestrom, P G et al Cell 19, 671–681 (1980)
113 Nagata, S, Mantei, N & Weissman, C Nature 287, 401–408 (1980)
114 Konkel, D A, Tilghman, S M & Leder, P Cell 15, 1125–1132 (1978)
115 Nishioka, Y & Leder, P Cell 18, 875–882 (1979)
116 Kitchingman, G, Lai, S & Westphal, H Proc natn Acad Sci USA 74, 4392–4395 (1977)
117 Fiers, W et al Nature 273, 113–120 (1978)
118 Soeda, E, Arrand, J R, Smolar, N, Walsh, J E & Griffin, B E Nature 283, 445–453 (1980)
119 Berk, A J & Sharp, P A Cell 14, 695–711 (1978)
120 Cold Spring Harb Symp quant Biol 44, Pt 2 (1979)
121 Stillman, B W, Lewis, J B, Chow, L T, Matthews, M B & Smart, J E Cell 23, 497–508 (1981)
122 Soeiro, R, Vaughn, M H, Warner, J R & Darnell, J E J Cell Biol 39, 112–118 (1968)
123 Lewin, B (ed) in Gene Expression 2nd edn, 748 (Wiley-Interscience, New York, 1980)
124 Ventateson, S, Nakazato, H, Kopp, D W & Edmonds, M Nucleic Acids Res 6, 1097–1110 (1979)
125 Davidson, E H & Britten, R J Science 204, 1052–1059 (1979)
126 Guyette, W A, Matuski, R J & Rosen, J M Cell 17, 1013–1023 (1979)
127 Greenberg, J R Nature 240, 102–104 (1972)
128 Palmiter, R J biol Chem 248, 8260 (1973)
129 Spector, D J, McGrogan, M & Raskas, H J J molec Biol 126, 395–414 (1978)
130 Aviv, H, Volloch, Z, Bastos, R & Levy, S Cell 84, 495–503 (1976)
131 Bastos, R N, Volloch, Z & Aviv, A J molec Biol 110, 191–203 (1977)
132 Volloch, V & Housman, D in Organization and Expression of Globin Genes (eds Stamatoyannopoulos, G & Nienhaus, A W) 251–268 (Liss, New York, 1981)
133 Chung, S, Landfear, S M, Blumberg, D D, Cohen, N S & Lodish, H F Cell 24, 785–797 (1981)
134 Brandhoest, B P Devl Biol 52, 310–317 (1976)
135 Rosenthal, E, Hunt, T & Roderman, J V Cell 20, 487–494 (1980)
136 McCormick, W & Penman, S J molec Biol 39, 315–333 (1969)
137 Kelley, P M & Schlessinger, M J Cell 15, 1277–1286 (1978)
138 Storti, R V, Scott, M P, Rich, A & Pardue, M L Cell 22, 825–834 (1980)
139 Fan, H & Penman, S J J molec Biol 50, 665–670 (1970)
140 Trachsel, H et al Proc natn Acad Sci USA 77, 770–774 (1980)
141 Skup, D, Zarbl, H & Millward, S J molec Biol 151, 35–56 (1981)
142 Fernandez-Munoz, R & Darnell, J E J Virol 18, 719–726 (1976)
143 Heywood, S M, Kennedy, D S & Bester, A J Proc natn Acad Sci USA 71, 2428–2431 (1974)
144 Roberts, J L & Herbert, E Proc natn Acad Sci USA 74, 4826–4830, 5300–5304 (1977)
145 Mains, R E & Epper, B A J biol Chem 251, 4115–4120 (1976)
146 Bach, M-L, Lacroute, F & Botstein, D Proc natn Acad Sci USA 76, 386–390 (1979)
147 Klar, A J S, Strathern, J N, Broach, J R & Hicks, J B Nature 289, 239–244 (1981)
148 Broach, J R J molec Biol 131, 41–53 (1979)
149 Hutchison, H T, Hartwell, L H & McLaughlin, C S J Bact 99, 807–814 (1969)
150 Hofer, E, Hofer-Warbinek, R & Darnell, J E Jr Cell (in the press)
151 Carlson, M & Botstein, D Cell 28, 145–154 (1982)

ARTICLES

Implications for trace gases and aerosols of large negative ion clusters in the stratosphere

F. Arnold, A. A. Viggiano & H. Schlager

Max-Planck-Institut fur Kernphysik, Heidelberg, FRG

Stratospheric ions are of considerable interest because they not only control atmospheric electrical properties but also may influence stratospheric trace gases and aerosols The first detailed in situ *composition measurements of the stratospheric heavy negative ion clusters are now reported, these measurements have interesting implications for trace gases and aerosols*

IN SITU composition measurements of stratospheric ions became technically possible only during recent years As found from the first *in situ* negative ion composition measurements[1,2], stratospheric sulphuric acid vapour has a strong influence on stratospheric negative ions Thus, because H_2SO_4 vapour is the most important condensable vapour in the stratosphere, and influences the formation of the stratospheric aerosol layer[3], negative ion composition measurements can also yield information on aerosols In particular $HSO_4^-(H_2SO_4)_b(HNO_3)_c$ ions have been found[4] to be abundant at altitudes above ~27 km, whereas $NO_3^-(HNO_3)_a$ ions are dominant below this height It has been proposed[1] that the former ions are formed from the latter ones by reactions involving H_2SO_4 vapour followed by H_2SO_4-association This suggestion has received strong support

0028-0836/82/220371—06$01 00

from studies of relevant ion reactions conducted by Viggiano *et al*[5] Attention has also been drawn[6] to the possibility of inferring stratospheric H_2SO_4-vapour concentrations from *in situ* negative ion composition measurements

Recent *in situ* negative ion composition measurements conducted by the Heidelberg[7–9] and Bruxelles[10] groups confirmed the original suggestions, but also provided strong evidence for the presence of very massive negative ion clusters in the region above 27 km In fact, it appears that most negative ions fall outside the mass range ($\leq$320 AMU), accessible to previous *in situ* ion composition measurements in which mass resolving modes of operation were used The so called 'total throughput modes' allowed detection of heavy negative ion clusters without yielding detailed mass information We have investigated large negative ion clusters in the stratosphere and their bearing on aerosol formation[11,12] Our detailed *in situ* measurements are given below

Negative ion composition measurements

The principle of the balloon-borne ion mass spectrometer used in the present studies has been discussed in detail elsewhere[1,4,12] and therefore will not be reviewed here Note, however, that the present instrument has an extended mass range covering 0–612 AMU, which was achieved by reducing the frequency of the R F -generator of the quadrupole mass filter from 2 0 to 1 4 MHz

The improved instrument was flown on a balloon during daytime on 17 October 1981 over southwestern France (geographical coordinates of the measuring site 44° N, 0° W) and reached a maximum altitude of 34 km The composition measurements of heavy negative ions were made between 32 and 34 km during the ascent and float portions of the balloon flight between 09 24 and 09 47 LT

Typical negative ion mass spectra obtained in this height region are shown in Fig 1 The mass resolution used was only

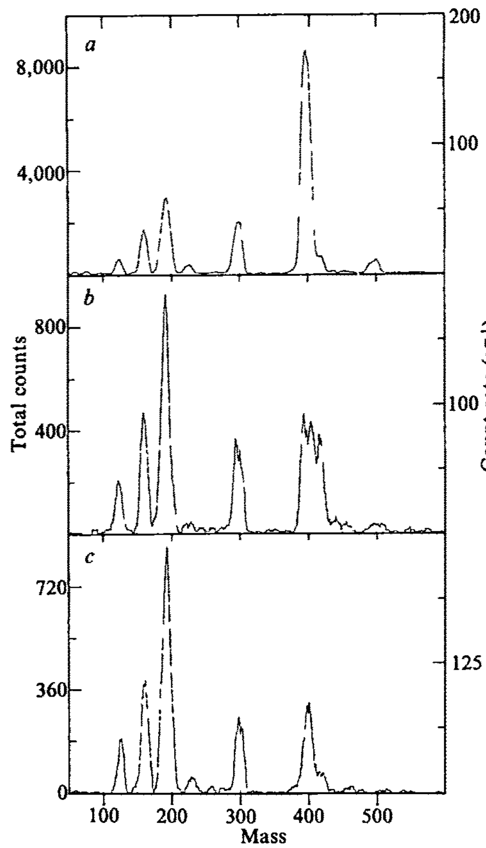


Fig 1 Negative ion mass spectra measured at altitudes of 34 (*a*), 32 9 (*b*) and 32 4 (*c*) km Total integration times were 1,000, 140 and 100 s A running window of 51 channels was used to improve statistics

moderate (average peak width at half maximum equal to about 14 AMU) to achieve a high sensitivity for ion detection, particularly at large masses Due to the large peak width, masking of small peaks may have occurred The estimated uncertainty for mass identification is about ±1 AMU large peaks and ±2 or 3 AMU for smaller peaks depending on the ion pulse statistics, which, in turn, depend on the integration time Due to a much larger integration time, the peaks of the spectrum obtained at 34 km (float altitude) have much better defined shapes

Results for the spectral region 0–320 AMU are rather similar to those obtained previously[1,2,4,7,10] Therefore, we will focus on the mass region 320–612 amu, which was not previously accessible

Clearly, this region contains major peaks, which is particularly striking in Fig 1*a*, where the dominant mass peak is 391±1 A compilation of the observed peaks is given in Table 1 together with fractional count rates and tentative ion identifications No attempt has been made to convert ion count rates to ion abundances as this conversion is complicated by cluster ion break-up that occurs during ion sampling As has been discussed previously[4], break-up should not drastically alter the ion composition at these heights However, cluster ions may have lost one weakly bonded ligand molecule during sampling

Negative ion chemistry

Table 1 reveals that the major heavy ions have HSO_4^- cores containing mainly H_2SO_4 ligands Size distributions for these ions are shown in Fig 2 The distributions are relatively flat up to $b=3$ (b is the number of ligands) and fall off steeply for larger b Most likely, these ions are formed by

$$NO_3^-(HNO_3)_a + H_2SO_4 \rightarrow HSO_4^-(HNO_3)_a + HNO_3 \quad (1)$$

followed by further displacement of HNO_3 by H_2SO_4 and finally H_2SO_4 association As b increases the thermal dissociation process

$$HSO_4^-(H_2SO_4)_b + M \rightarrow HSO_4^-(H_2SO_4)_{b-1} + H_2SO_4 + M \quad (2)$$

becomes important and leads to the observed fall-off of the size distribution

The predicted size distribution can be obtained from a steady-state treatment yielding

$$\frac{[HSO_4^-(H_2SO_4)_{b+1}]}{[HSO_4^-(H_2SO_4)_b]} = \left(1+\frac{\alpha n_+}{k_+[M][H_2SO_4]}+\frac{k[M]}{k_+[M][H_2SO_4]}\right]^{-1} \quad (3)$$

where k_+, k and α are the rate coefficients for ternary H_2SO_4-association, thermal dissociation and ion–ion recombination, respectively The quantities $[M]$ and n_+ are the total gas and positive ion concentrations Taking the phenomenological binary rate coefficient, $k_2=k_+[M]$, for H_2SO_4-association equal to 10^{-9} cm^3 s^{-1} (saturated ternary process) and a sulphuric acid vapour concentration $[H_2SO_4]=3\times10^6$ cm^{-3} (derived from the abundance ratio of measured HSO_4^- and NO_3^--ion clusters using the method described in ref 6) one finds that the second term on the right-hand side of equation (3) becomes $\ll 1$ Consequently, if H_2SO_4-association proceeds close to the collision rate, as expected, the ion abundance ratio (3) becomes $\simeq 1$ when thermal dissociation is negligible As seen from Fig 2, this seems approximately to be the case for b up to 2–3 When thermal dissociation becomes very efficient, the third term on the right-hand side of equation (3) becomes dominant and the predicted ion abundance ratio becomes $\ll 1$ This seems to be the case for $b \geq 3$

From Table 1, one sees that there are less abundant heavy ions containing HNO_3, H_2O and probably HSO_3 ligands While H_2O and HNO_3 ligands were expected, the presence of HSO_3 ligands comes as a surprise H_2O, HNO_3 and HSO_3 probably associate to the negative ion clusters and are lost by thermal dissociation, displacement by more strongly bonded ligands or

Table 1 Mass numbers of major negative ions as measured at 32 5, 32 9 and 34 0 km altitude

Mass	Fraction 006	Ion
62±2	0 006	NO_3^-
125±1	0 065	NO_3^- HNO_3
143±2	0 009	NO_3^- HNO_3 H_2O
160±1	0 140	HSO_4^- HNO_3
178±2	0 020	HSO_4^- HNO_3 H_2O
188±1	0 190	$NO_3^-(HNO_3)_2$
195±3	0 100	HSO_4^- H_2SO_4
223±1	0 018	$HSO_4^-(HNO_3)_2$
251±2	0 003	$NO_3^-(HNO_3)_3$
276±3	0 010	$HSO_4^-H_2SO_4$ HSO_3
293±1	0 120	$HSO_4^-(H_2SO_4)_2$
311±3	0 003	$HSO_4^-(H_2SO_4)_2H_2O$
374±3	0 003	$HSO_4^-(H_2SO_4)_2HSO_3$
391±2	0 140	$HSO_4^-(H_2SO_4)_3$
409±3	0 130	$HSO_4^-(H_2SO_4)_3H_2O$
427±3	0 022	$HSO_4^-(H_2SO_4)_3(H_2O)_2$
454±2	0 015	$HSO_4^-(H_2SO_4)_3HNO_3$
472±3	0 004	$HSO_4^-(H_2SO_4)_3HSO_3$
489±3	0 012	$HSO_4^-(H_2SO_4)_4$
507±3	0 015	$HSO_4^-(H_2SO_4)_4H_2O$
587±3	0 004	$HSO_4^-(H_2SO_4)_5$

Tentative ion identifications and fractional ion count rates for the 32 9 km spectrum are given

possibly by reactions The mass identity of the ligand we identify as HSO_3, however, does not exclude other possibilities For example, the mass of an HNO_3 and H_2O molecule added together (81 AMU) is the same as that of HSO_3 The relative abundances of the ions $HSO_4^-(H_2SO_4)_nH_2O$, $HSO_4^-(H_2SO_4)_nHNO_3$ and $HSO_4^-(H_2SO_4)_n$ 81, however, leads one to believe that mass 81 corresponds to only one ligand HSO_3 is then the most likely candidate for this ligand because it is the only acid of this mass to be expected in the atmosphere and acids are the preferred negative ion ligands[1,2 4]

Because both, H_2O and HNO_3 are relatively abundant stratospheric trace gases (approximate number densities are 10^{12} cm^{-3} and 10^8 cm^{-3}, respectively, at 30 km altitude), they associate rapidly to the negative ion clusters However, both molecular species probably do not bond very strongly to the larger ion clusters and, therefore, may undergo rapid thermal detachment or displacement by more strongly bonded ligands, such as H_2SO_4 The ligand bond strength seems to increase in the sequence H_2O, HNO_3, H_2SO_4 for the cluster ion sizes considered here

Measured abundance ratios for hydrate ions to their corresponding unhydrated forms are given in Table 2 Generally, the ratios decrease as the size of the unhydrated ion increases, in qualitative harmony with a simple charge–dipole interaction model However, for $HSO_4^-(H_2SO_4)_b$ clusters, the ratios increase as b becomes >2 This behaviour, which as yet has not been observed for cluster ions, probably indicates strong cooperative bonding In other words, ligand–ligand interaction seems to become important In view of the large heat of mixing for bulk phase H_2SO_4–H_2O mixtures, this finding may qualitatively be expected Quantitatively, however, it is surprising that cooperative effects become important already for relatively small ion cluster sizes

The probable detection of HSO_3-ligands indicates that HSO_3-vapour is present in the stratosphere and associates to the negative ion clusters

This view is supported by recent high pressure ion source studies made at our laboratory[13] It was found that besides $HSO_4^-(H_2SO_4)_b$, clusters of the type $HSO_4^-(H_2SO_4)_b$ HSO_3 were most abundant in a gas mixture of SO_2 and room air (mostly N_2, O_2, and H_2O) ionized by a gas discharge The HSO_3-ligand evidently bonds relatively strongly to $HSO_4^-(H_2SO_4)_b$ ion clusters This is also supported by collision-induced fragmentation studies of $HSO_4^-(H_2SO_4)_bHSO_3$ ions

The fate of stratospheric HSO_3-ligands is, however, uncertain They may detach thermally, become displaced by H_2SO_4, or undergo chemical reactions The latter processes may involve reactions with OH-radicals

$$HSO_4^-(H_2SO_4)_bHSO_3 + OH \rightarrow HSO_4^-(H_2SO_4)_b + SO_3 + H_2O \quad (4a)$$

$$\rightarrow HSO_4^-(H_2SO_4)_b \; H_2O + SO_3 \quad (4b)$$

$$\rightarrow HSO_4^-(H_2SO_4)_b + H_2SO_4 \quad (4c)$$

$$\rightarrow HSO_4^-(H_2SO_4)_{b+1} \quad (4d)$$

In process (4*d*) another ligand, for example H_2O, also contained in the precursor ion complex may become 'boiled off' carrying away the chemical energy released

Processes (4*a*–*d*) would represent 'ion catalysed reactions'[17]

Inferred HSO₃-abundance

Although the HSO_3-radical has not yet been detected in the stratosphere, its presence is expected It is thought[3 14,15] that HSO_3 arises from

$$SO_2 + OH + M \rightarrow HSO_3 + M \quad (5)$$

and is lost preferentially by

$$HSO_3 + OH \rightarrow SO_3 + M \quad (6)$$

whose rate coefficient is not known The SO_3 radical is then presumed to react with H_2O to form H_2SO_4 Within the framework of their complex photochemical–diffusive model, Turco *et al* have calculated the stratospheric HSO_3 abundance, yielding ~10^5 cm^{-3} at 30 km altitude

From the ion composition data reported above, a lower limit to the HSO_3 abundance can be inferred Assuming that $HSO_4^-(H_2SO_4)_bHSO_3$ ions are formed by HSO_3-association proceeding at the collision rate and that they are lost only by ion–ion recombination a steady-state treatment yields a lower limit of $[HSO_3] = (2–10) \times 10^4$ cm^{-3} When compared with this inferred value, the model calculation of Turco *et al* are in approximate agreement

As the model assumed reaction (6) to proceed at the collision rate (rate coefficient 10^{-11} cm^2 s^{-1}) the above agreement may imply that reaction (6) is in fact fast The HSO_3-lifetime compared with attachment to negative ions is ~10 times larger than the lifetime compared with reaction (6) if upper limits to the corresponding rate coefficients are used (ref 16, and personal communications from D Smith, N G Adams and E Alge, and P J Crutzen) Consequently, the HSO_3 abundance would be of the order of 10^6 cm^{-3} at 30 km if reaction (6) was unimportant and if the HSO_3-loss would proceed by ion attachment

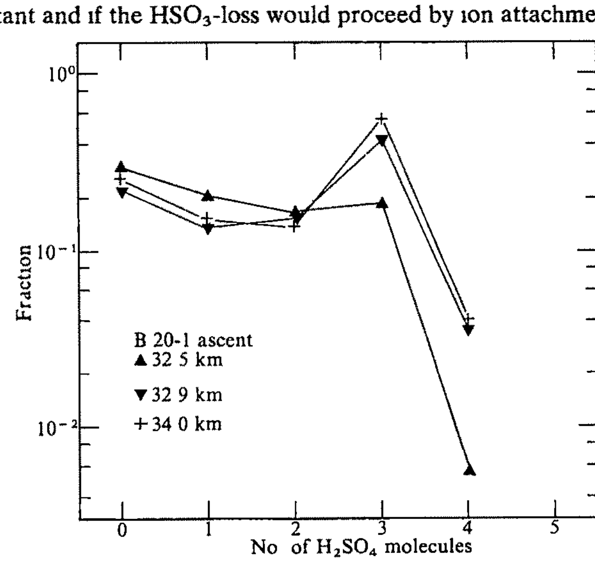


Fig. 2 Measured distributions of HSO_4^- clusters containing 0–5 H_2SO_4-ligands Count rates were normalized to the total count rates for HSO_4^- clusters

At lower altitudes, however, where the OH-abundance is lower and where the total ion concentration is larger, a significant fraction of HSO_3 may be lost by ion reactions Because, at these heights, $NO_3^-(HNO_3)_a$ ions are most abundant, another ion process involving HSO_3 may be considered It involves

$$NO_3^-(HNO_3)_a + HSO_3 \rightarrow HSO_4^-(HNO_3)_a + NO_2 \quad (7)$$

followed by recombination of $HSO_4^-(HNO_3)_a$ with a positive ion, mostly $H^+(CH_3CN)_l(H_2O)_m$ and $H^+(H_2O)_n$ possibly leading to the formation of an H_2SO_4 molecule Reaction (7) is exothermic[5 18 19] by ~40–50 kcal mol^{-1}

An important implication concerning analytical applications of *in situ* composition measurements of stratospheric negative ions is that the detection of HSO_4^- cores would not allow us unambiguously to infer H_2SO_4 abundances, thus the previously derived[6 8 9] H_2SO_4 abundances should be reported as H_2SO_4 and HSO_3 abundances below ~27 km Above this height, however, H_2SO_4 should be much more abundant than HSO_3 and the derived quantity should refer only to H_2SO_4 This conclusion is supported by the fact that above 27 km the negative ion clusters also contain H_2SO_4 ligands

Thermodynamic properties of ion clusters

The present ion composition data can also be used to infer unknown thermodynamic properties of large negative ion clusters

For this purpose, the concentrations of the clustering gas phase molecular species, H_2SO_4, HNO_3, HSO_3 and H_2O as well as the gas temperature have to be known Sulphuric acid concentrations were inferred independently from the negative ion composition (see above) whereas HNO_3-, HSO_3- and H_2O-concentrations were taken from model calculations The actual values used are $[H_2SO_4] = 3\times10^6\ cm^{-3}$, $[HNO_3] = 1\times10^8\ cm^{-3}$, $[HSO_3] = 1\times10^5\ cm^{-3}$, $[H_2O] = 1\times10^{12}\ cm^{-3}$ The gas temperature was measured by various radiosonde stations (CNES Meteorological Service, personal communication) in the vicinity of the balloon launch site on the day of our experiment In addition, the gas temperature can be inferred from the simultaneously measured positive ion composition with relatively high accuracy The resulting temperature for 32–34 km was consistently 233 K

The above concentrations, with the exception of $[HSO_3]$, are sufficiently large to allow for clustering time scales less than the ion–ion recombination lifetime (2,000 s) Consequently, quasi-equilibria for ion-clustering of these species should be established For HSO_3 clustering, this is not the case and therefore only lower limits to bond strengths can be inferred

First, we will examine H_2SO_4-clustering to $HSO_4^-(H_2SO_4)_b$ ions ($b = 2, 3, 4$) Equilibrium constants, K, can easily be inferred from the measured ion abundance ratios and the derived H_2SO_4 abundance The free energy difference, $\Delta G = -RT \ln K$, can then be evaluated using the measured gas temperature The resulting ΔG values are listed in Table 3 along with free enthalpy differences, $\Delta H = \Delta G + T\Delta S$, estimated by assuming a typical common entropy difference[20], ΔS, of 25 cal $mol^{-1}\ K^{-1}$

The inferred ΔG-values for H_2SO_4-clustering to $HSO_4^-(H_2SO_4)_b$ clusters with $b = 3$–4 are both about 13 kcal mol^{-1} which is comparable to the heat of vaporization for bulk phase H_2SO_4 (ref 21) (12 6 kcal mol^{-1} at 233 K) The inferred ΔG represents only a lower limit, because thermal dissociation of $HSO_4^-(H_2SO_4)_3$ does not seem to be sufficiently fast to allow for an equilibrium between $b = 2$ and 3 to be established (see Fig 2)

This result suggests that the properties of the observed heavy $HSO_4^-(H_2SO_4)_b$ clusters ($b = 4$ and 5) already exhibit a trend towards bulk phase properties Qualitatively, this may be expected as this trend is well known from laboratory studies[20 22 23] of cluster ion equilibria such as $H^+(H_2O)_n$, $NO_3^-(H_2O)_n$ and $Cl^-(SO_2)_n$ It was found that ΔH-values for clustering of a molecular species to ion approach the heat of vaporization of the molecular species as the number of ligands contained in the ion cluster increases Usually, only a few ligands (3–5) are needed to make the agreement better than 1 kcal mol^{-1}

Table 2 Ratios of measured hydrate ion abundances to their unhydrated precursors

Unhydrated species	Ratio
$NO_3^-(HNO_3)_1$	0 14
$HSO_4^-(HNO_3)_1$	0 14
$HSO_4^-(H_2SO_4)_2$	0 03
$HSO_4^-(H_2SO_4)_3$	0 92
$HSO_4^-(H_2SO_4)_4$	1 25
$HSO_4^-(H_2SO_4)_3H_2O$	0 17

Assuming that HSO_3 is indeed observed as a ligand, only lower limits to $-\Delta G$ can be inferred from the stratospheric ion composition data, as equilibrium abundances may not be reached and the concentration of HSO_3 is not easily derived These lower limits (Table 2) are comparable to the corresponding values for H_2SO_4-clustering, which suggests that the HSO_3-radical bonds strongly to HSO_4^- clusters In turn, this implies that the heat of vaporization of HSO_3 from an H_2SO_4–HSO_3 mixture should be >18 kcal mol^{-1}

Now we consider H_2O-clustering to negative ions Measured hydrate-ion distributions are given in Table 2 and corresponding ΔG and ΔH values are listed in Table 3 The extent of hydration first seems to decrease as the cluster size increases and that it increases for $HSO_4^-(H_2SO_4)_b$ clusters when $b \geq 2$ The former behaviour is known from laboratory studies[20 24] of negative ion hydrates containing H_2O-ligands only Unfortunately, no studies of mixed clusters containing H_2O- and H_2SO_4-ligands are available Qualitatively, a decrease of H_2O-bond energies with increasing cluster size should also be expected from a simple charge–dipole interaction model

Evidently, if the observed increase of ΔH for $b > 2$ is real, it reflects the occurrence of strong cooperative bonding between H_2O- and H_2SO_4-ligands When compared with the heat of vaporization for bulk phase H_2O (10 4 kcal mol^{-1}), our inferred ΔH-values for $HSO_4^-(H_2SO_4)_3(H_2O)_d$ ($d = 1, 2$) are markedly larger Qualitatively, this may be expected from bulk phase considerations, as the H_2SO_4–H_2O system has a large heat of mixing

Our inferred values for ΔH of 13–14 kcal mol^{-1} are much closer to the heat of vaporization of water from a bulk phase solution of 75% H_2SO_4 and 25% H_2O (13 kcal mol^{-1}), which is thought to be the composition of the stratospheric aerosol Additional strong support for the above interpretation comes from recent studies of negative ion hydration in our laboratory[13] where it was found that ΔG values for the hydration of the ion series $HSO_4^-(H_2SO_4)_b$, in fact, increase as b becomes >2

However, due to their small bond energies, negative ion hydrates may undergo substantial collisional dissociation while being sampled into the mass spectrometer This is supported by simulation experiments[13] in which it was found that in conditions similar to those for flight electric field-induced detachment of the weakly bonded H_2O molecules can be important But this effect should at most raise the $-\Delta G$ values by 1 kcal mol^{-1} (90% dissociation) and otherwise not significantly affect our interpretations

We now turn to HNO_3-clustering Here direct comparison with laboratory ΔH-data[20 25] can be made, at least for $NO_3^-(HNO_3)_2$ and $NO_3^-(HNO_3)_3$ and good agreement is found This supports our approach and suggests that no significant dissociation of these cluster ions occurs at the altitudes considered This does not conflict with the above remarks on H_2O dissociation because HNO_3 bonds more strongly to negative ions than does H_2O This conclusion is also supported by the observed large abundance of $H^+(H_2O)_4$ ions (bond energy

17 kcal mol^{-1}) which excludes significant dissociation of this ion Thus, for ligands with bond strengths >17 kcal mol^{-1} no significant dissociation seems to occur

Again, in the case of HNO_3 an increase in ΔG seems to occur at large cluster sizes ($HSO_4^-(H_2SO_4)_3HNO_3$) This may suggest that the heat of vaporization of HNO_3 from bulk phase H_2SO_4–HNO_3 is much higher than that for bulk phase HNO_3 which is only ~8 kcal mol^{-1} It seems, therefore, that the observed heavy HSO_4^--clusters exhibit properties which are similar to those of the bulk phase

Implications for aerosols

Our *in situ* composition measurements of stratospheric negative ions have important implications for stratospheric aerosols, concerning both their nature and their formation mechanism Because thermodynamic properties of large ion clusters seem to approach those of the corresponding bulk phase, information on the composition of aerosols may be obtained from studies of large ion clusters existing in the same environment The ion clusters studied in the present *in situ* measurements, however, are not very large and, therefore, their properties may still depart from those of the bulk phase Thus, the following discussion should be treated with caution

The largest negative ion clusters measured contain mostly H_2SO_4 and H_2O, and thus, they contain exactly the same molecular species which are thought to make up most of the stratospheric aerosol mass However, the ligand shell of the ion cluster $HSO_4^-(H_2SO_4)_3(H_2O)$, for example, contains much less water (up to about 8% by mass) than the aerosol solution droplets (about 25%) This discrepancy may be partly due to cluster ion fragmentation occurring during the measurement, but most of the discrepancy appears real

The detection of HSO_3 and the discovery that it bonds strongly to the large negative ion clusters suggests that the as yet unknown heat of vaporization of the HSO_3 radical is large, possibly comparable with that of H_2SO_4, which would imply that HSO_3 also undergoes efficient attachment to stratospheric aerosols

If HSO_3 is, indeed, a precursor of stratospheric H_2SO_4, this may imply that HSO_3 attaches to aerosol solution droplets and is then converted to H_2SO_4 by chemical reactions in the liquid phase Subsequently, H_2SO_4 may evaporate from the droplets and a condensation–evaporation equilibrium should be established according to temperature[29]

Thus, stratospheric aerosols may grow mostly by condensation of the HSO_3 radical and its hydrates rather than by H_2SO_4-condensation The possibility of HSO_3-nucleation and -condensation was originally discussed by Friend *et al*[26] their "radical nucleation" hypothesis being based on their laboratory studies of the SO_2-photooxidation

In view of the large estimated aerosol surface area density at heights below ~30 km gas phase HSO_3-radicals may therefore preferably attach to aerosols rather than attaching to ions or reacting with OH

At 20 km, for example, the estimated HSO_3-lifetimes against these processes are about 2×10^4 s, 7×10^4 s and 10^5 s At higher altitudes where the aerosol abundance decreases and where OH becomes more abundant, gas phase HSO_3 may preferably react with OH

Another implication of the present ion composition measurements concerns polyion nucleation It has been proposed[11] that the coagulation of large positive and negative ion clusters may lead to polyions or multi-ion complexes, which, in turn, may grow to the size of condensation nuclei The first step in polyion-formation is the formation of a stable ion pair by recombination of a negative and a positive cluster ion To prevent spontaneous neutralization and thereby destruction of the ion pair, the following criterion has to be fulfilled

$$S > IP - EA + E_C \qquad (8)$$

Table 3 Free energy- and enthalpy differences for ion clustering as inferred from stratospheric ion composition data (32 9 km altitude)

Ion	Ligands	$-\Delta G°$	$-\Delta H°$	$-\Delta_{lab}^{\circ}$
$HSO_4^-(H_2SO_4)_2$–H_2SO_4	3	14	19 8	
$HSO_4^-(H_2SO_4)_3$–H_2SO_4	4	13	18 5	
$HSO_4^-(H_2SO_4)_4$–H_2SO_4	5	13	18 9	
$HSO_4^-(H_2SO_4)_2$–HSO_3	3	>12	>18 0	
$NO_3^-HNO_3$–H_2O	2	7 1	12 9	
HSO_4^-–HNO_3–H_2O	2	7 2	13 0	
$HSO_4^-(H_2SO_4)_2$–H_2O	3	6 3	12 1	
$HSO_4^-(H_2SO_4)_3$–H_2O	4	8 0	13 8	
$HSO_4^-(H_2SO_4)_4$–H_2O	5	8 1	13 9	
$HSO_4^-(H_2SO_4)_3H_2O$–H_2O	5	7 2	13 0	
$NO_3^-HNO_3 - HNO_3$	2	12 6	18 4	18 3
$NO_3^-(HNO_3)_2$–HNO_3	3	10 2	16 0	16 1
$HSO_4^-HNO_3$–HNO_3	2	11 2	17 0	
$HSO_4^-(H_2SO_4)_2$–HNO_3	3	10 3	15 9	
$HSO_4^-(H_2SO_4)_3$–HNO_3	4	11 1	16 9	

Values are given in units of kcal mol^{-1} Laboratory data, if available, are also given for comparison For the derivation of ΔH as common ΔS of 25 cal mol^{-1} K^{-1} was assumed

where *IP* and *EA* are the ionization potential of the precursor neutral of the positive ion core and the electron affinity of the precursor neutral of the negative ion core E_C is the energy released on formation of a chemical bond, and S is the total solvation energy or, in other words, the total energy for all bonds between ligand molecules and the core ions

For $HSO_4^-(H_2SO_4)_b$, $H^+(H_2O)_n$ and $H^+(CH_3CN)_l(H_2O)_m$ ions, the most prominent ions around 32–34 km, the inequality (8) becomes

$$S > 13\ 6\ \text{eV} \qquad (9)$$

assuming that $H^+(CH_3CN)_l(H_2O)_m$ solvation energies are similar to those of $H^+(H_2O)_n$ This is supported by recent laboratory measurements[27] of the competitive solvation of a proton by CH_3CN and H_2O Here, a lower limit to $EA(HSO_4)$ = 4 5 eV as measured by Viggiano *et al*[5] is assumed E_0 has been set equal to the H–HSO_4 bond strength[19] While the total solvation energy for $H^+(H_2O)_n$ is known, that for $HSO_4^-(H_2SO_4)_b$ ions is not known for b equal to 1, 2, and 3 Assuming $\Delta H = 1, 2, 3$ to be the same as that deduced from the present data for $b = 4, 5$ (~20 kcal mol^{-1}) upper limits to the critical sizes of cluster ions having the potential to form stable ion pairs can be estimated The resulting minimum (n, b) pairs for critical clusters are $(2, 6)$, $(3, 5)$, $(4, 4)$, $(5, 3)$ and $(7, 2)$ Because ΔH values for $HSO_4^-(H_2SO_4)_b$ ($b = 1, 2, 3$) were underestimated and a lower limit to the electron affinity of HSO_4^- was used, the actual critical values may be smaller by about one ligand molecule

In view of the fact that $H^+(H_2O)_4$ and $H^+(H_2O)_5$ are the most abundant proton hydrates around 32–34 km, it seems that ion–ion recombination of the newly observed large HSO_4^--clusters may mainly lead to stable ion pair formation Taking fractional ion abundances equal to measured fractional count rates, the fraction of ion–ion recombinations able to form stable ion-pairs is estimated to be about 0 1–0 7 around 32–34 km

Thus, at the altitudes considered, the abundant negative ion clusters seem to grow just large enough to allow for stable ion-pair formation on mutual recombination with the dominant positive ion clusters

At still lower altitudes, $NO_3^-(HNO_3)_a$ and positive ion clusters may again ultimately become large enough to allow for high polyion formation rates However, in this region, small polyions, due to the large aerosol surface area density, may rapidly attach to aerosols before they reach nuclei sizes

Because negative ion clusters are largest around 30–35 km and because the size of positive ion clusters decreases as height increases the region around 30 km seems to be particularly favourable for polyion formation Around this altitude, there may exist a layer of seed particles, which on mixing into the region below 30 km, where HSO_3 or H_2SO_4 become highly supersaturated, may initiate nuclei formation

Conclusions

We have shown that *in situ* composition measurements of stratospheric ions can provide valuable information on thermodynamic properties of large ion clusters and aerosols In addition, they yield new information on condensable vapours like H_2SO_4 and possibly also HSO_3 Thus, *in situ* ion composition measurements strongly contribute to our understanding of stratospheric trace gas and aerosol processes.

Substantial progress in the new field of stratospheric ion composition measurements can be expected when it becomes possible to lower ion detection limits and to reduce collisional dissociation occurring during ion sampling Thus, it may be possible to detect even larger ion clusters having low abundances and bond energies

We thank R Turco, J Friend, P Crutzen and D Ehhalt for helpful discussions We also thank the technical staff of the Cosmophysics Department of the Max-Planck-Institute for Nuclear Physics, particularly K Oberfrank, W Thron, D Dorflinger, H Sauer, B Preissler and K Bechberger for their encouragement without which this project would not have been possible Part of the project was funded by the Bundesministerium für Forschung und Technologie through Gesellschaft fur Strahlen- und Umweltforschung

Received 29 December 1981, accepted 31 March 1982

1 Arnold, F & Henschen, G *Nature* **257**, 521 (1978)
2 Arnold, F, Fabian, R, Ferguson, E E & Joos, W *Planet Space Sci* **29**, 195 (1981)
3 Turco, R P, Hamill, P, Toon, O B, Whitten, R C & Kiang, C S *J atmos Sci* **86**, 699 (1979)
4 Viggiano, A A & Arnold, F *Planet Space Sci* **29**, 895 (1981)
5 Viggiano, A A, Perry, R A, Albritton, D L, Ferguson, E E & Fehsenfeld, F C *J geophys Res* **85**, 4551 (1980)
6 Arnold, F & Fabian, R *Nature* **283**, 55 (1980)
7 McCrumb, J L & Arnold, F *Nature* **294**, 136 (1981)
8 Arnold, F, Fabian, R & Joos, W *Geophys Res Lett* **8**, 293 (1981)
9 Viggiano, A A & Arnold, F *Geophys Res Lett* **8**, 583 (1981)
10 Arijs, E, Nevejans, D, Frederick, P & Ingels, J *J geophys Res Lett* **8**, 121 (1981)
11 Arnold, F *Nature* **284**, 610 (1980)
12 Arnold, F, Henschen, G & Ferguson, E E *Planet Space Sci* **29**, 185 (1981)
13 Glebe, W & Arnold, F *Proc Spring Meet Arbeitsgemschaft fur Extraterrstriche Forschung*, Koln (in the press)
14 Castleman, A W J Jr, Davis, R E, Munkelwitz, H R, Tang, I N & Wood, W P *Int J chem Kinet Symp* **1**, 629 (1975)
15 Moortgat, G K. & Junge, C E *Pure appl Geophys* **115**, 759 (1977)
16 Rosen, J M & Hofmann, D J *J geophys Res* **86**, 7399 (1981)
17 Rowe, B R, Viggiano, A A, Fehsenfeld, F C, Fahey, D W & Ferguson, E E *J Chem Phys* **76**, 742 (1982)
18 Stull, D R & Prophet, H (eds), *JANAF Thermodynamical Tables* (Nat Stand Ref Data Ser, National Bureau of Standards Washington DC, 1971)
19 Bensen, W W *Chem Rev* **78**, 23 (1978)
20 Lee, N, Keessee, R G & Castleman, A W Jr *J chem Phys* **72**, 1089 (1980)
21 Ayers, C P, Gillett, R W & Gras, J L *Geophys Res Lett* **7**, 433 (1980)
22 Kebarle, P *A Rev phys Chem* **28**, 445 (1977)
23 Keessee, R G, Lee, N & Castleman, A W Jr *J chem Phys* **13**, 2195 (1980)
24 Keesssee, R G., Lee, N & Castleman, A W Jr *J Am chem Soc* **101**, 2559 (1979)
25 Davidson, J A, Fehsenfeld, F C & Howard, C J *Int J Mass Spectrom Ion Phys* **9**, 17 (1977)
26 Friend, J P, Barnes, R A & Vista, R M *J phys Chem* **84**, 2423 (1980)
27 Bohringer, H & Arnold, F *Nature* **290**, 321 (1981)
28 Hamill, P, Turco, R P, Toon, O B, Kiang, C S & Whitten, R C *J atmos Sci* (in the press)

A computational model of binocular depth perception

J. E. W. Mayhew* & H. C. Longuet-Higgins†

* Department of Psychology, University of Sheffield, Sheffield S10 2TN, UK
† Laboratory of Experimental Psychology, University of Sussex, Brighton, Sussex BN1 9QG, UK

We point out that the horizontal disparities between a pair of retinal images are inadequate for computing the three-dimensional structure of a scene unless supplemented by independent information about the distance and direction of the fixation point We suggest that this supplementary information is derived not from non-visual sources, but from the vertical disparities of a few non-meridional image points This hypothesis is shown to account quantitatively for Ogle's induced effect—the marked distortion of a scene by a vertically magnifying lens placed in front of one eye

THE binocular perception of depth may be regarded as proceeding in two stages the establishment of a point-by-point correspondence between the left and right retinal images to extract disparity information, and the interpretation of the disparities to yield a three-dimensional percept The position of an image point on the left retina can be represented by its plane projective coordinates (x, y), if (x', y') are the coordinates of the corresponding image point on the right retina, then the difference $x'-x$ is described as the horizontal disparity between the two image points, and $y'-y$ as the vertical disparity The 'correspondence problem'—that of computing the disparities—has been fully discussed in the literature[1-5] and we shall not consider it further Our present concern is the 'interpretation problem'—that of using the disparities for computing the three-dimensional structure of the scene.

As vertical disparities produced by local depth variations (in contrast to those produced by asymmetrical convergence) are commonly much smaller than horizontal ones[1,8]—and necessarily vanish on the horizontal meridian—it is often supposed that only the horizontal disparities are relevant to the binocular perception of depth There are two difficulties with this hypothesis First, the horizontal disparities supply insufficient information, even in principle, for computing the absolute distances and directions of a set of visible points, they must be supplemented by independent information about the distance and direction of the fixation point, but non-visual estimates of these 'viewing parameters' are notoriously unreliable There is, however, a more serious difficulty the binocular perception of a three-dimensional scene is profoundly altered by distorting one of the retinal images in the *vertical* dimension If a cylindrical lens which induces a small vertical magnification is placed in front of one eye, then a visually textured surface in the fronto-parallel plane will acquire a pronounced tilt, as if it had been rotated about a vertical axis, although the horizontal disparities and the non-visual cues to fixation distance and direction are quite unaffected by the introduction of the lens

0028-0836/82/220376—03$01 00

The question therefore arises could the vertical disparities supply the information about the viewing parameters that is required for deriving the structure of the scene from the horizontal disparities? The answer, as we shall see, is in the affirmative, and the hypothesis that the viewing parameters are derived from the vertical disparities rather than from non-visual information sources accounts quantitatively for the above phenomenon, which Ogle has described as the 'induced effect'[1]

It has recently been pointed out[6,7] that a binocular system which takes account of the vertical as well as the horizontal image coordinates can in principle solve the interpretation problem without recourse to extra-retinal sources of information, and also[8] that there exists a remarkably simple approximate method of deriving the position and orientation of a visually textured plane from the horizontal and vertical disparities of a small number of texture elements lying in the plane Figure 1 illustrates the geometry of this situation O and O′ are the nodal points of the two eyes, and the observer is looking—not quite straight ahead—at the plane $Z = PX + QY + R$, where the Z axis joins the mid-point of O O′ to the fixation point $(0, 0, R)$ and makes an angle g with the forward direction The capital letters refer to a coordinate system with the origin at the midpoint of the line joining the optical centres of the eyes If the fixation distance R is large enough compared with the interocular distance I, then one can neglect, in the disparities, terms of the second order in I/R It then becomes possible to use the algebra that Longuet-Higgins and Prazdny[14] used in their analysis of the visual perception of motion, the problem being isomorphous with that presented by the stereoscopic viewing of an object at a moderate distance If $\sin(g) = s$ and $\cos(g) = c$, one obtains, in this way,

$$H(x, y) = x' - x = [(Pc + s)x + Qcy + (c - Ps)x^2 - Qsxy]I/R \quad (1)$$

$$V(x, y) = y' - y = [sy + (c - Ps)xy - Qsy^2]I/R \quad (2)$$

These equations simplify if g is small enough for s and c to be replaced by g and 1 respectively, and for Pg and Qg to be neglected, in these conditions

$$H(x, y) = x' - x = [(P + g)x + Qy + x^2]I/R \quad (3)$$

$$V(x, y) = y' - y = (gy + xy)I/R \quad (4)$$

These equations immediately suggest a way of computing the unknown parameters P, Q, R and g The equation for V implies that a plot of V/y against x gives a line of slope I/R and intercept gI/R, so that the values of R and g can be obtained from the vertical disparities of just two non-meridional points with different x coordinates (that is, points with $y \neq 0$ and $x_1 \neq x_2$) Explicitly

$$R = I/V_{xy} \qquad g = V_y/V_{xy} \quad (5)$$

where V_y and V_{xy} denote the coefficients of y and xy in the expression for V Likewise, if H_x and H_y denote the coefficients of x and y in H, it follows from equation (3) that

$$P + g = H_x R/I, \qquad Q = H_y R/I \quad (6)$$

Substituting from equation (5) into equation (6), we infer that

$$P = (H_x - V_y)/V_{xy}, \qquad Q = H_y/V_{xy} \quad (7)$$

In short, the viewing parameters R and g can be directly computed from the vertical disparities alone, once they are known the geometrical parameters P and Q may be computed from either equation (6) or equation (7), in which the viewing parameters do not appear explicitly A computer simulation implementing equations (5) and (7) has shown that the small-angle approximations underlying these equations are indeed adequate for a wide range of reasonable surface orientations, viewing distances and angles of gaze

When a vertically magnifying lens is placed in front of one eye, the two retinal images become mutually incompatible, in the sense of ceasing to admit of a consistent three-dimensional interpretation It is therefore quite surprising that the observer should obtain a clear three-dimensional percept, albeit a distorted one The induced effect therefore offers a challenge to theories of stereopsis Would a 'seeing machine' which was based on a particular theory be subject to the induced effect, or would it arrive at a different interpretation of the two images, or fail to reach any interpretation at all? For example, machines based on the theories described in refs 6 and 7 would not be subject to the induced effect and therefore may be excluded as models of the human visual process Similarly, a full implementation of equations (1) and (2) would not in general show the induced effect, whereas a machine implementing the approximate solutions [equations (3) and (4)] apparently interprets the two images in a similar way to that reported by human subjects We shall now show that a system which computes P and Q directly from equation (7) will be subject to an induced effect of exactly the same nature and magnitude as that investigated by Ogle

Suppose that a lens of horizontal magnification λ and vertical magnification μ is placed over the right eye Then the new retinal coordinates for that eye will be

$$\hat{x} = \lambda x', \qquad \hat{y} = \mu y' \quad (8)$$

so that the horizontal and vertical disparities will be

$$\hat{H} = \hat{x} - x = \lambda[(P + g)x + Qy + x^2]I/R + (\lambda - 1)x \quad (9)$$

$$\hat{Y} = \hat{y} - y = \mu(gy + xy)I/R + (\mu - 1)y \quad (10)$$

The new values of V_x and V_{xy} will therefore be

$$\hat{V}_x = \mu gI/R + (\mu - 1) = \mu(V_y + 1) - 1 \quad (11)$$

$$\hat{V}_{xy} = \mu I/R = \mu Y_{xy} \quad (12)$$

and the new values of H_x and H_y will be

$$\hat{H}_x = \lambda(P + g)I/R + (\lambda - 1) = \lambda(H_x + 1) - 1 \quad (13)$$

$$\hat{H}_y = \lambda QI/R = \lambda H_y \quad (14)$$

The apparent values of P, Q and R will therefore be

$$\hat{P} = (\hat{H}_x - \hat{V}_y)/\hat{V}_{xy} = [\lambda(H_x + 1) - \mu(V_y + 1)]/\mu V_{xy} \quad (15)$$

$$\hat{Q} = \hat{H}_y/V_{xy} = \lambda H_y/\mu V_{xy} = \lambda Q/\mu \quad (16)$$

$$\hat{R} = I/\hat{V}_{xy} = I/\mu V_{xy} = R/\mu \quad (17)$$

If λ and μ are both $\simeq 1$, the values of Q and R are not much affected, but with P it is an entirely different matter To see

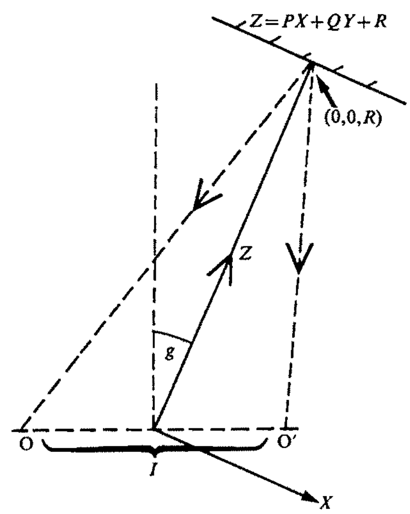


Fig. 1 The coordinate system has origin at the midpoint of the line joining the optical centres of the eyes (O and O′) The line joining this point to the point of fixation $(0, 0, R)$ defines the Z axis, the X axis lies in the plane containing the line OO′ and fixation point, and the line normal to this plane is the Y axis The angle of gaze g, is the angle OO′ makes with the X axis The analysis assumes that the eyes are free to rotate around their X and Y axes but not their Z axes

this, suppose that P is actually zero, as it will be for a fronto-parallel plane Then, according to equation (7), H_x and V_y must be equal, and so the apparent value of P, given by equation (15), will be

$$\hat{P} = (\lambda - \mu)(1 - V_y)/\mu V_{xy} \tag{18}$$

As V_y is of order I/R, and λ and μ are close to 1, we may neglect V_y in the numerator and infer that

$$\hat{P} \simeq (\lambda - \mu)/V_{xy} = (\lambda - \mu)R/I \tag{19}$$

Equation (19), with μ set equal to 1, becomes Ogle's own expression for the apparent tilt produced by a horizontally magnifying lens—an effect which can be understood in purely geometrical terms and which he therefore named the 'geometric effect' Various predictions follow, and all are confirmed by Ogle's observations

(i) The apparent tilt is proportional to the difference between the horizontal and vertical magnification, so that for a given value of R a vertical magnification induces an apparent tilt equal and opposite to that produced by an equivalent horizontal magnification (The induced effect actually falls off at high magnification, this is not predicted by equation (19), but is presumably connected with the fact that $\hat{V}_y/\hat{V}_{xy}$ then implies an implausibly large value for the angle of gaze)

(ii) An ordinary spherical lens, with $\lambda = \mu$, induces no tilt

(iii) When $\mu = 1$ and λ near 1 the induced tilt is proportional both to $\lambda - 1$ and to R At a viewing distance $R = 40$ cm, Ogle found that the angle of tilt increased by 3–3 5° for every 1% increase in magnification With an interocular distance $I =$ 6 5 cm, equation (19) predicts a value for this ratio of

$$(180/\pi) \times (40/6\ 5) \times (1/100) = 3\ 53° \text{ per \% magnification}$$

In his experiments, Ogle actually used a null method of estimating the induced tilt of a vertical plane the subject had to tilt he plane himself (in the opposite direction) until the apparent slope P became zero, that is, until $\hat{H}_x = \hat{V}_y$ This relationship entails that

$$\lambda(H_x + 1) = \mu(V_y + 1) \tag{20}$$

or

$$(\lambda - \mu) = \lambda H_x - \mu V_y = [\lambda(P + g) - \mu g]I/R \tag{21}$$

Remembering that $\lambda = 1$ and $\mu = 1$, we get

$$\mu - \lambda \simeq PI/R \tag{22}$$

showing that the actual value of P required to make $P = 0$ is

$$P = (\mu - \lambda)R/I \tag{23}$$

This is, indeed, equal and opposite to the apparent tilt of a fronto-parallel plane under identical viewing conditions

Discussion

It has commonly been believed (see ref 9 for a review) that only the horizontal disparities convey useful information about the distances of objects, and only qualitative information at that Vertical disparities, which arise when an object is nearer one eye than the other, have been regarded merely as something that the visual system must be capable of allowing for in fusing the two images into a unitary percept Our approach to the problem, from an artificial intelligence standpoint, has led to a computational solution of the interpretation problem which accounts quantitatively for the induced effect Various explanations of this effect have been proposed, both by Ogle and by others[1,2 9–12], but none, we feel, follows so naturally from a theory of normal stereoscopic depth perception as the explanation proposed above (the limitations of a recent theory[12] which attempts to explain the induced effect without using vertical disparities have been criticised elsewhere[11]) Ogle's explanation, like ours, agrees quantitatively with the psychophysical magnitude of the induced effect, but requires the use of the vertical disparities on the vertical meridian to compute the compensatory rotation of the Veith–Muller circle necessary to maintain the correct egocentric localization (this corresponds to the direction of gaze in our expressions) However, Ogle's theory fails to exploit the vertical disparity information at other retinal locations and thus does not solve for the other viewing parameter, distance, which remains as an unknown scaling variable in his expression Note that the retinal eccentricity of the location at which the disparity information is measured is of fundamental importance in our expressions relating horizontal and vertical disparity to the viewing and scene parameters [equations (1)–(4)] Without the terms derived solely from the retinal eccentricity, the system of equations would be homogeneous and without a unique solution The emphasis on a coordinate system derived from concern with corresponding points—the horopter and the Veith–Muller circle—may have masked the importance of these terms from Ogle's consideration, and may explain why, when he was so close to a complete solution, he failed to find it

As our theory is formally complete, it predicts that the induced effect will result from any stimulus which supplies misleading cues about the viewing parameters—the distance and direction of the fixation point Although the vertical disparities from any two retinal locations could be used to solve for the viewing parameters, and thus the computation could be essentially local to a particular region of retina, the results of the computation have global implications which can provide powerful constraints and checks on its consistency In contrast to the interpretation of horizontal disparities where the results in one region of the retina exert little constraint on the interpretation at another region, the interpretation of the vertical disparities should give the same result everywhere There are obvious advantages in making viewing system parameters globally available, and so one might expect that the induced effect would act on the whole visual field even though vertical disparity information could only be derived from a particular region of it Ogle[1] reports an experiment confirming this expectation Vertical disparities from one part of the visual field changed the apparent depths of vertical rods (which of course cannot provide vertical disparity information) in accord with the predictions of the induced effect If the stimulus situation presents incompatible vertical disparity information at different parts of the visual field or, alternatively, vertical disparity information whose interpretation results in grossly implausible viewing system parameters, one might expect some difficulty in perceiving the induced effect This may explain why there may be some difficulty in producing the induced effect without the aid of a cylindrical lens[12 13]

Finally, it is of interest that Ogle reports excellent agreement in estimates of stereoacuity derived from experiments investigating the induced effect and other psychophysical procedures, which correspond to a sensitivity of about 0 2% magnification between the two eyes' images The constraints that this degree of resolution imposes on the implementation of the processes described above are currently the subject of computational and psychophysical experimentation

Received 13 October 1981 accepted 26 March 1982

1 Ogle, K N *Binocular Vision* (Saunders, Philadelphia, 1950)
2 Julesz, R *Foundations of Cyclopean Perception* (University of Chicago Press, 1971)
3 Nelson, J *J theor Biol* **49**, 1–88 (1975)
4 Marr, D & Poggio, T *Proc R Soc* **R204**, 301–328 (1979)
5 Mayhew, J E W & Frisby, J P *Art Intelligence* **23**
6 Longuet-Higgins, H C *Nature* **293**, 133–134 (1981)
7 Longuet-Higgins, H C *Perception* (in the press)
8 Mayhew, J E W *Perception* (in the press)
9 Foley, J M *Psychol Rev* **87**, 411–434 (1980)
10 Nelson, J *J theor Biol* **66**, 203–266 (1977)
11 Mayhew, J E W & Frisby, J P *Vision Res* (in the press)
12 Arditi, A, Kaufman, L & Movshon, J A *Vision Res* **21**, 755–764 (1981)
13 Westheimer, G *Invest Opthal Visual Sci* **17**, 545–551 (1978)
14 Longuet-Higgins, H C & Prazdny, K *Proc R Soc* **B208**, 385–397 (1980)

LETTERS TO NATURE

Could primordial black holes be the source of the cosmic ray antiprotons?

Michael S. Turner

Astronomy and Astrophysics Center, The University of Chicago, Chicago, Illinois 60637, USA

Antiprotons ($\bar{p}$) are expected to be found in the cosmic rays at the level of $\sim 10^{-4}$ relative to protons (p) as they should be produced by cosmic-ray proton collisions with the interstellar medium[1-3]. The recent observations of Buffington *et al*[4]. are puzzling, however, because the total $\bar{p}$ flux observed is ~ 10 times larger than predicted, and the shape of the $\bar{p}$ spectrum differs greatly from that predicted. We explore here the possibility that primordial black holes (PBHs) of mass $\sim 10^{13}$–10^{15} g which evaporated after decoupling ($t \gtrsim 10^{14}$ s) are the primary source of the cosmic ray $\bar{p}$s. The PBH scenario predicts a universal $\bar{p}$/p ratio of $\sim 10^{-4}$–10^{-8}.

Cosmic-ray protons are generally believed to traverse ~ 5 g cm^{-2} during their life in the Galaxy Cosmic-ray proton collisions with the interstellar gas are therefore predicted to produce an $\bar{p}$ flux of $\bar{p}/p \simeq 2\times 10^{-4}$ for $\bar{p}$ kinetic energies above a few GeVs The predicted $\bar{p}$/p ratio falls very steeply for $\bar{p}$ energies below about 1 GeV (refs 1–3) The recent measurements of Buffington *et al*[2] in the energy range 130–320 MeV indicate a $\bar{p}$ flux of $\sim 2\times 10^{-5}$ cm^{-2} sr^{-1} s^{-1} GeV^{-1}, which is several orders of magnitude larger than the flux expected if the $\bar{p}$s are cosmic ray secondaries The three measurements made so far span the energy range from ~ 100 MeV to ~ 10 GeV, and are consistent with a constant $\bar{p}$/p ratio of $\sim 3\times 10^{-4}$ The integrated $\bar{p}$ flux is about a factor of 10 larger than is predicted

There are three possible explanations of the discrepancy (1) The measurements are wrong The results for the energy range of a few to 10 GeV are only a factor of two to three higher than the predicted flux[5,6] Buffington *et al*'s[4] measurement shows greater divergence (2) The $\bar{p}$s observed are indeed secondary particles, but the standard picture of cosmic ray propagation must be modified For example, some of the cosmic ray protons would have to traverse significantly more grammage than ~ 5 g cm^{-2} thus producing more $\bar{p}$s[7-9], and a stochastic mechanism[11] to decelerate the $\bar{p}$s would be required to explain the high flux of low-energy $\bar{p}$s Buffington *et al*[4] observe (3) The $\bar{p}$s detected are primary particles originating from the big bang, or from regions of antimatter If their origin is the big bang, then there is the problem of storing them so that they escape annihilation[12], and then subsequently injecting them into the Galaxy If the observed $\bar{p}$s originate from antigalaxies[13], then it is difficult to understand why the $^4\overline{\rm He}/^4$He ratio is $<2\ 2\times 10^{-5}$ (95% confidence, ref 4), rather than being equal to the $\bar{p}$/p ratio $\simeq 3\times 10^{-4}$

Here I discuss three models in which evaporating PBHs of mass $\sim 10^{13}$–10^{15} g radiate $\bar{p}$s sufficiently late in the history of the Universe that the $\bar{p}$s avoid annihilation, leading to a universal abundance of $\bar{p}/p \sim 10^{-4}$–10^{-8} The initial spectrum of PBHs required to explain the observations is consistent with previous constraints on the spectrum of PBHs[14], and could also lead to the production of a significant baryon asymmetry[15] (much earlier in the history of the Universe, $t\sim 10^{-35}$ s) Because the PBHs must have a temperature of $\gtrsim 1$ GeV when they produce the $\bar{p}$s, antinuclei will not be produced in the process, thus explaining the absence of antinuclei

First, consider the question of producing $\bar{p}$s that do annihilate with the ambient protons in the Universe The $\bar{p}$ annihilation rate $\Gamma_{\rm ann}$ is given by

$$A6\Gamma_{\rm ann} = n_{\rm p}(\sigma v)_{\rm ann} \qquad (1)$$

The number density of protons, $n_{\rm p}$, is just $(n_{\rm p}/n_\gamma)\times n_\gamma \simeq (3\times 10^{-10})\times(2\zeta(3)T^3/\pi^2)$, where $n_{\rm p}/n_\gamma \simeq (3\text{–}5)\times 10^{-10}$ (J Yang *et al*, in preparation) and the second factor is the number density of photons at temperature T, $\zeta(3)\simeq 1\ 202$ The annihilation cross-section times the relative velocity $(\sigma v)_{\rm ann} \simeq m_\pi^{-2}$ Annihilations will not occur at a significant rate if $\Gamma_{\rm ann} < H \equiv \dot{R}/R = 1\ 66\ g_*^{1/2}T^2/m_{\rm pl}$ (the expansion rate) Here $\hbar = c = k_{\rm B} = 1$, $m_{\rm pl} = G^{-1/2} \simeq 10^{19}$ GeV and g_* is the total number of effective degrees of freedom of all the relativistic species at a temperature T The condition $\Gamma_{\rm ann} < H$ requires

$$T \lesssim 10^{10}(\pi^2/2\zeta(3))m_\pi^2/m_{\rm pl} \sim 0\ 1\ {\rm eV} \qquad (2)$$

This temperature roughly corresponds to that of decoupling ($T\sim 1{,}000$K), and to a time $t\sim 10^{14}$ s Essentially all the $\bar{p}$s produced after $t\sim 10^{14}$ s will avoid subsequent annihilation, while $\bar{p}$s produced earlier will be annihilated (If $n_{\rm p}/n_\gamma \gtrsim (3-5)\times 10^{-10}$, then the Universe will be matter-dominated when the annihilations cease and the analysis above must be modified If, however, the Universe is nucleon-dominated now, then the time at which annihilations cease is independent of $n_{\rm p}/n_\gamma$ If the Universe is not nucleon-dominated and $\Omega h^2 \simeq 1$, $t\simeq 3\times 10^{12}$ s)

Hawking[16] showed that a black hole of mass m should radiate like a blackbody with temperature

$$T \simeq 0\ 1 m_{\rm p1}(m_{\rm p1}/m) \qquad (3)$$

and evaporate in a time[17]

$$\tau \simeq m_{\rm p1}^{-1}(m/m_{\rm p1})^3\,(100/g_*) \qquad (4)$$

where g_* is the number of species with mass less than T Of course, this process is only important for black holes of mass $m \ll 1M_\odot$ Such black holes must be primordial (produced from density perturbations in the early Universe, see, for example, ref 13), as known astrophysical processes can only produce black holes of mass $M \gtrsim$ few $M_\odot$

Equation (4) shows that PBHs are ideal candidates for producing $\bar{p}$s after $t\sim 10^{14}$s, as they have a built-in clock PBHs of mass $m \gtrsim 10^{18}m_{\rm p1} \sim 10^{13}$ g will evaporate after $t\sim 10^{14}$ s A PBH of mass $m\sim 10^{18}\ m_{\rm p1}$ has a temperature of ~ 1 GeV As it radiates, its mass decreases and its temperature increases If we assume that for $T \gtrsim 1$ GeV about one-tenth of the particles the PBH radiates are $\bar{p}$s[14], and the typical $\bar{p}$ momentum is $\sim T$, then the spectrum of $\bar{p}$s produced is

$$dn_{\bar{p}} \simeq 2\times 10^{36}p^{-3}\,dp \qquad (5)$$

where the momentum p is measured in GeV A PBH more massive than $10^{18}\ m_{\rm p1}$ will not emit $\bar{p}$s until its mass becomes less than $\sim 10^{18}\ m_{\rm p1}$, and then it will emit the same spectrum as a PBH of that mass, given by equation (5)

Following Carr[14], we parameterize the initial PBH spectrum as

$$\frac{dn(m)}{n_\gamma} = \zeta\left(\frac{m}{m_{\rm p1}}\right)^{-\alpha}\frac{dm}{m_{\rm p1}} \qquad (6)$$

where $dn(m)/n_\gamma$ is the number of PBHs per photon in the mass interval $m\to m+dm$, ζ is the normalization ($\simeq$fraction of the mass density of the universe in PBHs at $t_{\rm p1}\sim 10^{-43}$ s) and α is the index of the spectrum The most stringent constraint on ζ and α results from the insistence that PBHs of mass $\sim 10^{15}$ g do not contribute excessively to the diffuse γ-ray background[14],

$$\zeta \lesssim 10^{20\alpha - 75} \qquad (7)$$

Of course, ζ must also be <1 [In equations (6) and (7) it is assumed that the number of PBHs per photon has remained constant (neglecting evaporation) This relies on the assumption that the expansion of the Universe has been approximately isentropic, so that the number of photons in the Universe has not increased significantly]

0028–0836/82/220379—03$01 00

Summing the contribution of all PBHs more massive than $\sim 10^{13}$g and taking into account the subsequent cosmological redshift of the $\bar{p}$ momenta, we find that, at present,

$$\frac{dn_{\bar{p}}}{n_\gamma} \simeq 10^{55\ 5-19\ 5\alpha}\zeta \mathrm{KE}^{-3/4-\alpha/4}\, d\,\mathrm{KE} \tag{8}$$

for $\bar{p}$ kinetic energies (measured in GeV) in the range 1 GeV$\geqslant$ KE$\geqslant 10^{-6}$ GeV The lowest-energy $\bar{p}$s are those that were emitted with momentum $\sim$1 GeV by 10^{13} g PBHs just after decoupling, at present they have KE$\sim 10^{-6}$ GeV ($v \sim 10^{-3}c$) For relativistic $\bar{p}$s ($E \gtrsim$ few GeV),

$$\frac{dn_{\bar{p}}}{n_\gamma} \simeq 10^{55\ 5-19\ 5\alpha}\zeta E^{-3}\, dE \tag{9}$$

None of the $\bar{p}$s produced by PBH evaporations would have initially collapsed into galaxies with the baryons, both because the $\bar{p}$s are emitted during the epoch of galaxy formation ($z \leqslant 10^3$) and since they are relativistic when they are emitted they are moving too fast to be captured in galaxies or protogalaxies The entire spectrum, equations (8) and (9), is primarily due to the $\sim 10^{13}$g PBHs which evaporated shortly after decoupling

The universal abundance of $\bar{p}$s relative to photons is obtained by integrating over the spectrum

$$\frac{n_{\bar{p}}}{n_\gamma} \simeq 10^{54\ 5-18\alpha}\zeta \leqslant 10^{-13} \tag{10}$$

and it is dominated by the low-energy $\bar{p}$s ($v \simeq 10^{-3}c$) The upper limit in equation (10) is obtained by using the maximum value of ζ consistent with the limit derived from the diffuse γ-ray background ($\zeta = 1$, $\alpha = 3\ 75$) Equation (10) translates into a universal $\bar{p}$/p ratio $\leqslant 10^{-4}$

Antiprotons are not annihilating rapidly enough now to deplete their abundance significantly, however, the π^0s produced in $\bar{p}$p annihilations contribute to the diffuse γ-ray background when they decay To compute the annihilation rate one must use the nonrelativistic annihilation cross-section, $(\sigma v)_{\mathrm{ann}} \simeq 7\times 10^{-17}\ \mathrm{cm}^3\ \mathrm{s}^{-1}\ (c/v)$ (ref 18) The contribution to the γ-ray background today is

$$I \simeq [(\bar{p}/p)/10^{-4}]3\times 10^{-5}\ \mathrm{cm}^{-2}\ \mathrm{sr}^{-1}\ \mathrm{s}^{-1} \tag{11}$$

The flux observed at $E_\gamma \sim 100$ MeV is about $3\times 10^{-5}\ \mathrm{cm}^{-2}\ \mathrm{sr}^{-1}\ \mathrm{s}^{-1}$ (ref 19) The γ-ray background thus places an upper bound on the universal $\bar{p}$/p ratio of $\sim 10^{-4}$ $(v/10^{-3}c)$, regardless of their source (assuming that they are non-relativistic, see also ref 12)

From the abundance of $\bar{p}$s due to PBH evaporations [equations (8) and (9)], one must calculate the resulting galactic cosmic ray $\bar{p}$ flux For reference, the observed $\bar{p}$ flux corresponds to a local antiproton number density of

$$n_{\bar{p}}|_{\mathrm{galaxy}} \simeq 10^{-14}\ \mathrm{cm}^{-3} \tag{12}$$

essentially all in the energy range 0 1–1 GeV If the Galaxy is taken to be a sphere of radius $\sim$30 kpc, and the $\bar{p}$ number density constant, there are about 10^{55} $\bar{p}$ in the Galaxy today The escape time for cosmic rays of energy greater than several GeV is $\sim 10^7$ yr, however, lower-energy cosmic rays may be trapped for longer In any case, the annihilation time scale for an $\bar{p}$ in the Galaxy is $\sim 10^8$ yr Thus a reasonable estimate for the lifetime of an $\bar{p}$ in the Galaxy is 10^7–10^8 yr, and so $\bar{p}$s of energy 0 1–1 0 GeV must be replenished at the rate of

$$\frac{dN_{\bar{p}}}{dt} \simeq 10^{39}\text{–}10^{40}\ \mathrm{s}^{-1} \tag{13}$$

This is the rate at which the ambient intergalactic $\bar{p}$s must be injected into the Galaxy

There are three possible injection models (1) low-energy ($v \simeq 10^{-3}c$)$\bar{p}$s are captured by the Galaxy and then accelerated up to $\sim$0 1–1 0 GeV, (2) $\bar{p}$s in the energy range 0 1–1 0 GeV are captured by the Galaxy, (3) PBHs of mass $\sim 10^{15}$ g which are in our Galaxy and are evaporating today supply the required flux of $\bar{p}$s All three models are viable, and have associated uncertainties and difficulties

In the case of model (1), the total flux of low-energy $\bar{p}$s ($v \simeq 10^{-3}c$) on the Galaxy is given by

$$\frac{dN_{\bar{p}}}{dt} = 10^{111-18\alpha}\zeta\ \mathrm{s}^{-1} \leqslant 10^{44}\ \mathrm{s}^{-1} \tag{14}$$

To supply the required number of $\bar{p}$s in the energy range 0 1–1 0 GeV, the efficiency of capture and acceleration must be $>10^{-4}$

In model (2), the total flux of $\bar{p}$s in the energy range 0 1–1 0 GeV on the Galaxy is given by

$$\frac{dN_{\bar{p}}}{dt} = 10^{115-19\ 5\alpha}\zeta\ \mathrm{s}^{-1} \leqslant 10^{43}\ \mathrm{s}^{-1} \tag{15}$$

In this case to supply the required numbers of $\bar{p}$s in the range 0 1–1 0 GeV, the probability that a given $\bar{p}$ with energy $\sim$0 1–1 0 GeV incident on the Galaxy enters the galactic magnetosphere must be $\gtrsim 10^{-4}$ In both models (1) and (2) the $\bar{p}$s must scatter off inhomogeneities in the galactic magnetic field in order to become galactic cosmic rays This process is not well understood, and remains the biggest uncertainty in the estimates for models (1) and (2) In both models the $\bar{p}$s come primarily from $\sim 10^{13}$ g PBHs which evaporated at $t \sim 10^{14}$ s

In the case of model (3), although the bulk of the $\bar{p}$s in the intergalactic medium are produced by $\sim 10^{13}$ g PBHs which evaporate just after decoupling, $\sim 10^{15}$ g PBHs which are evaporating now are the primary source of galactic $\bar{p}$s If we assume that these PBHs collapsed with the baryons during galaxy formation, then their local number density is enhanced by a factor of $\sim 10^6$ relative to equation (6) The $\bar{p}$s such PBHs radiate now have energies 0(1 GeV), and are already inside the Galaxy The total flux of $\bar{p}$s from $\sim 10^{15}$ g PBHs in the Galaxy is

$$\frac{dN_{\bar{p}}}{dt} \sim 10^{117-20\alpha}\zeta\ \mathrm{s}^{-1} \leqslant 10^{42}\ \mathrm{s}^{-1} \tag{16}$$

For the maximum value of ζ, the total flux is 10^2–10^3 larger than required Thus one could improve Carr's constraint[14] on ζ slightly, $\zeta \leqslant 10^{-77+20\alpha}$ This model requires the number density of 10^{15} g PBH to be within $\sim 10^2$–10^3 of the maximum allowed number, it has the advantage that the $\bar{p}$s do not have to penetrate the galactic magnetosphere to become cosmic rays, hence avoiding the uncertainties involved with this process

It is therefore possible for the evaporation of $\sim 10^{13}$–10^{15} g PBHs to produce the observed, but unexplained, large flux of $\sim$0 1–1 0 GeV $\bar{p}$s The number density of PBHs must be within a factor of $\sim 10^4$ of the upper limit allowed by the γ-ray background constraint[14] The mass spectrum of PBHs can have an index α in the range 2–3 75 In the first and second models α must be $\sim$3 75 unless the injection of $\bar{p}$s into the Galaxy is very efficient If PBH evaporations are responsible for the cosmic ray $\bar{p}$s, then there should be a universal $\bar{p}$/p ratio of $\sim 10^{-4}$–10^{-8} in the intergalactic medium In two of the models discussed, intergalactic $\bar{p}$s must penetrate the galactic magnetosphere to become cosmic rays In the third model, the cosmic ray $\bar{p}$s are produced by $\sim 10^{15}$ g PBHs within the Galaxy which are evaporating today None of the PBH models is compelling, but they are plausible, and perhaps even intriguing

After completing this paper, I realised that Kiraly *et al*[20] have also suggested PBHs as a source of the cosmic ray antiprotons The model they discuss is similar to my model, in which PBHs evaporating in the Galaxy today are responsible for the cosmic ray $\bar{p}$s

I thank Tom Gaisser for introducing me to this problem and for beneficial discussions, and Gary Steigman for useful comments This work was supported in part by the DOE (ACO2-80ER10773) at University of Chicago and by NSF grant PHY 77-27084 at University of California, Santa Barbara

Received 10 November 1981, accepted 17 February 1982

1 Gaisser, T K & Levy, E H Phys Rev D10, 1731 (1974)
2 Stephens, S A Astrophys Space Sci 76, 87 (1981)
3 Szabelski, J, Wdowczyk, J & Wolfendale, A W Nature 285, 386 (1980)
4 Buffington, A & Schindler, S M Astrophys J Lett 247, L105 (1981)
5 Golden, R L et al Phys Rev Lett 43, 1196 (1979)
6 Bogomolov, E A et al Proc 16th int Cosmic Ray Conf, Kyoto, Vol 1, 33 (1979)
7 Rasmussen, I L & Peters, B Nature 258, 412 (1975)
8 Peters, B & Westergaard, N J Astrophys Space Sci 48, 21 (1977)
9 Stephens, S A Astrophys Space Sci 76, 87 (1981)
10 Steigman, G Astrophys J Lett 217, L131 (1977)
11 Szabelski, J, Wdowczyk, J & Wolfendale, A W Nature 285, 386 (1980)
12 Steigman, G A Rev Astr Astrophys 14, 339 (1976)
13 Stecker, F W, Protheroe, R & Kazanas, D Proc 17th int Cosmic Ray Conf, Paris (in the press)
14 Carr, B J Astrophys J 201, 1 (1975), 206, 8 (1976)
15 Turner, M S Phys Lett 89B, 155 (1979)
16 Hawking, S W Nature 248, 30 (1974)
17 Page, D N Phys Rev D13, 198 (1975)
18 Morgan, D L Jr & Hughes, V W Phys Rev D2, 1389 (1970)
19 Fitchel, C E et al Astrophys J 198, 163 (1975)
20 Kiraly, P, Szabelski, J, Wdowczyk, J & Wolfendale, A W Nature 293, 120 (1981)

Isotopically distinguishable carbon phases in the Allende meteorite

P. K. Swart, M. M. Grady & C. T. Pillinger

Planetary Sciences Unit, Department of Earth Sciences, University of Cambridge, Cambridge CB2 3EQ, UK

The stepwise oxidation of the Allende meteorite reported here has shown that there are at least three isotopically distinct carbon phases, each combustible over different temperature ranges. The temperature of combustion is dependent both on grain size of the fragments studied and/or on the nature of the carbonaceous material and its availability to the oxygen. Thus we attempt here to correlate the isotopic signatures recognized with phases whose existence has been postulated to explain noble gas systematics[1,2] and to provide some insight into the nature and location of these phases. Most of the carbon ($\delta^{13}C$= −20‰), possibly a highly cross-linked polymer, exists predominantly in the matrix and has been reported by others[3] to be the host of most of the normal planetary noble gases. Carbon protected in mineral grains may be isotopically heavier at −13 to −7‰, but is still more easily oxidized than graphite when exposed by HF/HCl demineralization. A very easily oxidized phase, as yet unidentified, has a $\delta^{13}C$ value typically ~−23 to −27‰.

Carbon isotopic analyses of carbon-rich meteorites (CI, CM2, C3 and type 3 unequilibrated ordinary chondrites), apart from H_3PO_4-released fractions[4], have revealed a very uninspiring range of values In fact, after removal of obviously separable minor components, such as carbonate and soluble organics, some of which may be contaminants, bulk carbon usually has an isotopic composition of between −5 and −28‰ and shows no coherent trend relative to meteorite group or petrological type[5] The bulk of the carbon in chondrites is largely uncharacterized and only operationally defined Thus, being insoluble in mineral acids and organic solvents, it can only be some allotrope or polymer of carbon The root cause of the unprepossessing isotopic compositions may be that bulk carbon is an unresolved mixture in such meteorites With material of several different isotopic compositions present in varying proportions, bulk values would be at best confused Recognition of several carbonaceous noble gas host-phases within a single meteorite[1,2] apparently endorses such an interpretation Obviously a method capable of further resolving mixtures within the bulk carbon would possibly allow isotope signatures to be recognized

The Allende C3V meteorite has a reported carbon isotope value of between −16 4 and −19 5‰ (refs 6–8) Much effort has been made recently to separate carbon-rich phases, thought to be among the hosts of anomalous noble gases, from the meteorite[1,2] In one such study, a separate (AMII)[9] was combusted and carbon isotopic analyses of the CO_2 produced reported by Ott *et al*[2] The data showed increasing heavy isotope abundance with temperature At the time, this change was thought to be a result of kinetic isotope effects and hence dismissed as of no significance[2] Recently, however, we have obtained evidence which suggests that such an interpretation may have been premature We performed a series of stepwise oxidations using various bulk sieve fractions of the Allende meteorite (>50μm, <50μm and <25μm) as well as a sample from which silicate had been partially removed by HF/HCl treatment, in an effort to bridge the gap between fully demineralized and bulk samples The method involved is similar to that used by others[2,3] for stepwise oxidations and involves the combustion of each sample for 30 min in an atmosphere of O_2(>20 mbar) derived from the decomposition of CuO Gases evolved are frozen into a variable temperature cryogenic trap and excess O_2 adsorbed onto a CuO/Cu furnace Carbon dioxide produced is separated from SO_2 and measured with a capacitance manometer Isotopic measurements were performed using a VG Micromass 602-E and data reported in ‰ relative to PDB To evaluate the extent of kinetic isotope effects during stepwise combustion, several carbonaceous substances were analysed using the same method

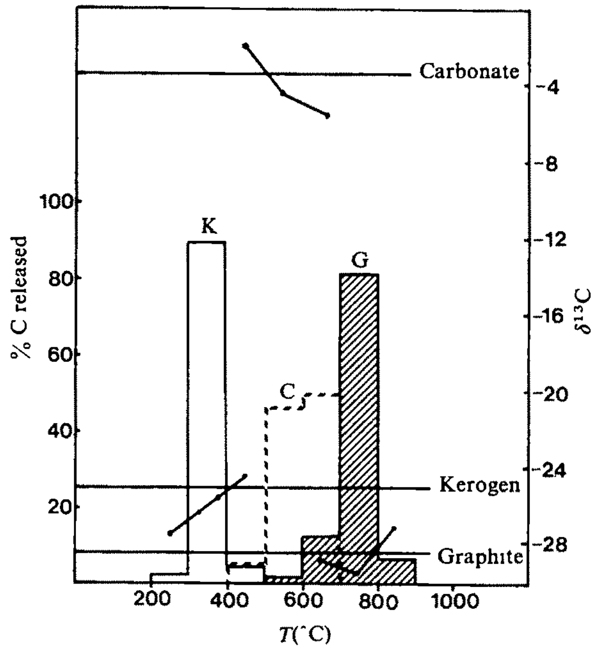


Fig. 1 Stepwise oxidation of three standard carbonaceous substances Kerogen (type I kerogen extracted from Upper Jurassic Kimmeridge Clay), PSU-1 carbonate (sparry calcite standard), and graphite (NBS-21) The $\delta^{13}C$ values obtained by a one-step total extraction are shown as horizontal lines

Stepwise carbon release patterns and isotopic compositions for the standard and meteorite samples are shown in Figs 1 and 2 For the <50 μm fraction, carbon yield and isotopic values, calculated by summing the individual temperature steps, agree well with measurements made during conventional extraction procedures, bulk measurements have not been made on the other fractions Unfortunately, the higher temperature extractions of the HF/HCl-treated sample were contaminated with fluorinated species and no isotopic data were obtained However, on the basis of previous results[2], we assume that the 500–800 °C fractions would have contained slightly heavier carbon We discount the interpretation that an increase in heavy isotope abundance with temperature of oxidation, seen in all cases, merely reflects kinetic isotope fractionation for several reasons (1) None of the standard materials exhibited a variation >±1 5‰ about their mean $\delta^{13}C$ values, although weighted averages of the $\delta^{13}C$ obtained in a stepwise fashion are comparable with that derived from a single step combustion (2) The spread of Allende isotopic measurements is 15–20‰ irrespective of grain size and is independent of whether any silicate has been removed from the sample The release pattern of the

0028–0836/82/220381—03$01 00

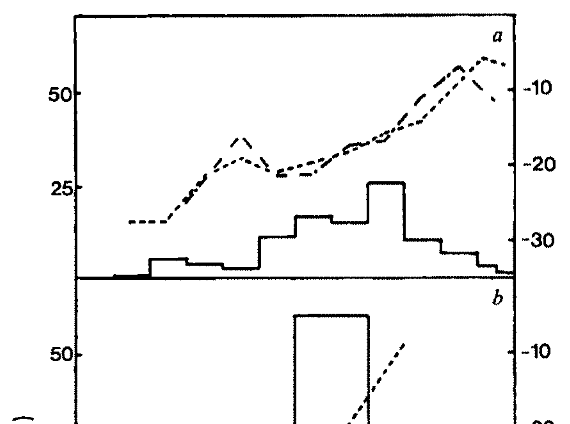


Fig. 2 Per cent carbon release and $\delta^{13}C$ composition from various size fractions and treatments of Allende *a*, >50 μm bulk duplicate analysis, $\Sigma\delta^{13}C = -17.2$, -17.1‰, $\Sigma C = 0.21\%$, only one carbon release profile is shown, *b*, <50 μm bulk, $\Sigma\delta^{13}C = -18.8$‰, $\Sigma C = 0.27\%$, *c* <25 μm bulk, $\Sigma\delta^{13}C = -18.9$‰, $\Sigma C = 0.37\%$, *d*, partially demineralized by HF/HCl, $\Sigma\delta^{13}C = -17.4$‰, $\Sigma C = 1.89\%$

carbon from the sample is, however, governed by the physical condition of the sample, being easier to combust in the absence of minerals and much more difficult to release from the >50 μm fraction The opportunity for kinetic isotope effects in the latter sample, which was analysed in duplicate, is obviously greater but clearly not excessively manifest (3) None of the undemineralized meteorite samples show a regular and continuous increase in $\delta^{13}C$ with temperature In the samples of >50 μm and <50 μm, troughs occur in the isotope release pattern over the 500–700 °C interval The <25 μm sample exhibited two distinct plateaus corresponding to peaks in the carbon release (4) Stepwise combustion applied to other meteorite samples gives rise to complex release profiles which cannot be explained by simple kinetic fractionation[10]

We suggest that for Allende, different forms of carbon are being oxidized or decomposed as temperature increases Undoubtedly, some kinetic isotopic fractionation occurs, but this is probably within a few per mil, based on evidence from our stepwise combustion of individual carbonaceous substances The profile obtained for the <25 μm fraction provides the best estimate of the isotopic compositions of the carbon phases (Fig 2) We have estimated the isotopic compositions on two criteria first, the carbon shows sharp release profiles and second for the most part these profiles appear to correspond with plateaus in the carbon isotope ratio The initial low temperature release (below 400 °C) shows a rapid increase in isotopic ratio to the first plateau, perhaps as a result of mixing of two or more phases having only slightly different oxidation temperatures The mid-temperature oxidizable phase, which contributes the bulk of the carbon is released over the interval 450–600 °C and has a mean $\delta^{13}C$ of ~−20 3‰ The high temperature carbon in the <25 μm fraction is liberated above 600 °C and has the heaviest carbon encountered in this size fraction and averages out at −13 4‰ Higher $\delta^{13}C$ values may be encountered in coarser size fractions (Fig 2) although these could reflect an enhanced kinetic isotope effect

In the duplicate >50 μm sized fractions of Allende, a reduction in the carbon isotopic ratio between 500 and 600 °C was observed Such a pattern of isotopic release could be explained as the existence of a further, possibly low abundance, carbonaceous phase possessing a very negative isotopic composition During stepwise combustion of <4 μm Allende matrix samples[11], the $\delta^{15}N$ plummets to −53‰ for the 1 28 p p m of nitrogen liberated during the 490 °C step and correlates strongly with the release of a heavy xenon component suggesting the addition of an isotope of nucleosynthetic origin[11] The admix of nearly pure ^{14}N of presolar origin into Solar System materials has previously been discussed by several workers[12–14] There may be no *a priori* significance between compatible release of nuclear components and a possible low $\delta^{13}C$, but the coincidence merits further study

The carbon release and isotope patterns of the various fractions studied allow us to draw some conclusions regarding the nature and location of the phases believed to exist First, we note that the mid-temperature phase (<25 μm nomenclature, Fig 2*c*) which constitutes most of the carbon in the meteorite is of variable temperature release when other fractions are considered Coarse sieve fractions do not release the phase completely until 900 °C, but our partially demineralized sample and the AMII fraction of Ott *et al*[2] suggest that in its free form this material combusts between 300 and 600 °C Such a reaction temperature is above the value found for a type I terrestrial kerogen, but well below that observed for NBS graphite and lends credence to the interpretation that the major carbon phase in Allende is a highly cross-linked, possibly aromatic, polymer, although some allotrope of carbon such as carbyne or amorphous carbon cannot be ruled out We thus concur with the interpretation of Simmonds *et al*[15] that carbon in Allende is in a macromolecular form unlike graphite We are unable to comment on the oxidation properties of "the interwoven ribbon-shaped packets of graphite layer plains" of Smith and Buseck[16] or poorly crystalline graphite[17] Although direct comparisons are difficult because of sample differences, the mid-temperature phase is most probably that discussed extensively by Frick and Pepin in relation to noble gases and nitrogen[9] The mid-temperature carbon phases must be associated with a portion of the meteorite which can be easily crushed and available for oxidation, the evidence seems to point strongly towards the fine-grained olivine matrix which has been shown to have carbonaceous surface deposits[17,18] Second, the high temperature carbon phase of the <25 μm fraction does not show a pronounced temperature variation when other sieved undemineralized fractions are considered Thus the heaviest carbon in the sample probably resides within the silicate minerals and corresponds to the so-called 'protected phase' of Frick and Pepin[3] Once the silicate minerals are partially or completely removed, the carbon becomes oxidizable below 700 °C suggesting that it is neither a highly crystalline nor ordered graphite Third, the very low temperature phase released below 400 °C in all undemineralized samples is clearly available to oxidation Although precautions were taken, it cannot be ruled out that a proportion of terrestrial contamination is involved Finally, we perceive, as previously reported by Chang *et al*[6] an increase in carbon content and a slight shift

towards lighter carbon as particle size decreases. These trends may be accounted for by an additional availability of the low and medium temperature phases and concurrent with their occurrence in fine-grained, easily crushed matrix.

We conclude that stepwise combustion of bulk and demineralized samples provides an easy and relatively contamination-free method of identifying the carbon isotopic composition of phases within Allende. The use of the technique in conjunction with noble gas and other volatile element analyses and its application to model materials should prove highly informative.

We thank the SERC for support, Professor E. Anders for stimulating our interest in stepwise combustion, the members of the Planetary Sciences Unit, Cambridge for helpful discussions and Mr I. Gilmour for the kerogen sample.

Received 17 February; accepted 22 April 1982.

1. Anders, E. *Proc. R. Soc.*, **A374**, 207–238 (1981).
2. Ott, U., Mack, R. & Chang, S. *Geochim. cosmochim. Acta* **45**, 1751–1788 (1981).
3. Frick, U. & Pepin, R. O. *Earth planet. Sci. Lett.* **56**, 45–63 (1981).
4. Clayton, R. N. *Science* **140**, 192–193 (1963).
5. Grady, M. M., Swart, P. K. & Pillinger, C. T. *Lunar planet. Sci.* **13**, 277–278 (1982).
6. Chang, S., Mack, R. & Lennon, K. *Lunar planet. Sci.* **9**, 157–159 (1978).
7. Mühle, K., Schnikow, B. L., Stiehl, G. & Winkler, I. *Chem. Erde.* **39**, 27–36 (1980).
8. Grady, M. M., Swart, P. K. & Pillinger, C. T. *Meteoritics* **16**, 319 (1981).
9. Frick, U. & Pepin, R. O. *Lunar planet. Sci.* **11**, 303–305 (1980).
10. Swart, P. K., Grady, M. M. & Pillinger, C. T. *Lunar planet. Sci.* **13**, 788–789 (1982).
11. Frick, U. & Pepin, R. O. *Earth planet. Sci. Lett.* **56**, 64–81 (1981).
12. Kung, C. C. & Clayton, R. N. *Earth planet. Sci. Lett.* **38**, 421–435 (1978).
13. Thiemans, M. H. & Clayton, R. N. *Earth planet. Sci. Lett.* **55**, 363–369 (1981).
14. Geiss, J. & Bochsler, P. *Geochim. cosmochim. Acta* **46**, 529–548 (1982).
15. Simmonds, P. G., Bauman, A. J., Bollin, E. M., Gelpi, E. & Oró, J. *Proc. natn. Acad. Sci. U.S.A.* **64**, 1027–1034 (1967).
16. Smith, P. P. K. & Buseck, P. R. *Science* **212**, 322–324 (1981).
17. Bauman, A. J., Devaney, J. R. & Bollin, E. M. *Nature* **241**, 264–267 (1973).
18. Green, H. W., Radcliffe, S. V. & Heuer, A. H. *Science* **172**, 936–939 (1971).

Rainfall acidity in northern Britain

D. Fowler*, J. N. Cape*, I. D. Leith*, I. S. Paterson†, J. W. Kinnaird† & I. A. Nicholson†

* Institute of Terrestrial Ecology, Bush Estate, Penicuik, Midlothian EH26 0QB, UK
† Institute of Terrestrial Ecology, Hill of Brathens, Glassel, Banchory, Kincardineshire AB3 4BY, UK

The acidity of rain has attracted considerable international interest during the past decade. An area of enhanced rainfall acidity with its focus in the Low Countries of north-west Europe was identified by the European air chemistry study[1] for the period 1955–68, and widespread concern arose as a result of observations that fish populations were decreasing in areas of Scandinavia subject to acid rain[2]. We report here measurements of acidity and related anion concentrations in rain collected at 16 sites in northern Britain during the period 1978–80. Estimates are provided of annual deposition of acidity in rain, the proportion of this acidity attributable to sulphuric and nitric acids, and an example of the influence of local concentrations of pollutant gases on rainfall acidity in cities.

The term 'acid rain' refers to rain more acid than water in equilibrium with atmospheric concentrations of carbon dioxide (~*p*H 5.6, 2.5 μEquiv. $H^+ l^{-1}$). Although several factors are involved, the areas in Scandinavia with significantly altered fish populations are subject to rainfall acidities of 30–60 μEquiv. $H^+ l^{-1}$ (*p*H 4.5–4.2). Fish mortality in acid water is caused mainly by problems with body salt regulation, the increased solubility of aluminium compounds in acid water being strongly implicated in these effects[3]. Evidence for links between acid precipitation and fish mortality is not restricted to Scandinavia, as similar effects have been reported for eastern North America[4] and Scotland[5].

In the United Kingdom there have been few detailed and systematic observations of rain chemistry in recent years, despite early interest in the subject[6]. Gorham, in 1958, showed that areas of north-west England received appreciable inputs of pollutants, including acidity, from industrial areas to the south-east[7]. Ten years later, Stevenson reported concentrations of all the major ions in rain except hydrogen for eight sites in Britain[8]. In 1978 and 1979, Martin and Barber showed that rain in central and eastern England was 'acidic', with annual average hydrogen ion concentrations of 40–80 μEquiv. $H^+ l^{-1}$ (*p*H 4.4–4.1)[9,10].

At each of the 16 sites (Fig. 1) rain was collected using a 20 cm diameter Pyrex glass funnel mounted 2 m above ground, each funnel surrounded by a ring of spikes to discourage birds from perching on the rim. Pyrex glass has been shown to influence concentrations of sodium and potassium ions through leaching, and may underestimate acidity[11] but the surface properties are likely to remain more consistent over prolonged periods in the field than plastic funnels which deteriorate with time. Rain water flowed into darkened polypropylene bottles through a small piece of nylon gauze inserted into the neck of the funnel to prevent the collection of insects and larger debris. Although preservatives were not added, there was only very occasional evidence of algal growth occurring between the monthly collections of samples, which were sent by post to the laboratory where they were stored at 4 °C before being analysed. There remains the possibility of chemical changes occurring in the sample through biological activity during the intervals between collection. However, Galloway and Likens[11] showed that there were no significant changes in concentrations of the major ions, except chloride, over a 7 month storage period at 21 °C for a sample at *p*H 4. A continually exposed funnel rather than an automatic wet-only sampler was used as there were problems of reliability and contamination with the automatic samplers available in 1977. The contribution of dry deposition

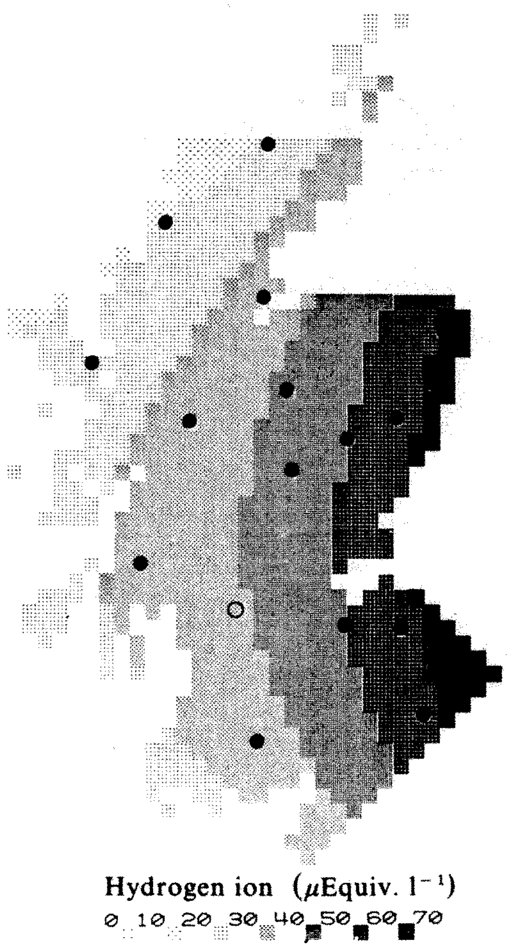


Fig. 1 Weighted mean acidity of rain for the period 1978–80 in northern Britain from equation (1) (●, rural sites; ○, urban site).

0028–0836/82/220383—04$01.00

Table 1 Site details and acidities, predicted* and measured, 1978–80

Site name	Map ref. (O.S.)	Altitude (m)	Measured (M) acidity (μEquiv. l^{-1})	Predicted (P) acidity (μEquiv. l^{-1})	P−M (μEquiv. l^{-1})
Rural sites					
Bettyhill	NC 708 624	30	25	21	−4
Inverpolly	NC 075 145	30	21	19	−2
Nigg	NH 795 738	5	27	34	+7
Broadford	NG 628 247	15	21	24	+3
Forest of Deer	NJ 976 510	100	60	61	+1
Achnagoichan	NH 914 082	300	37	42	+5
Banchory	NO 680 988	140	59	57	−2
Lochnagar	NO 274 858	600	52	50	−2
Torlundy	NN 147 773	30	33	32	−1
Faskally	NN 919 600	150	50	44	−6
Lephinmore	NS 005 915	300	37	32	−5
Bush	NT 246 638	185	48	49	+1
Whiteadder†	NT 663 633	230	57	58	+1
Redesdale	NY 833 954	260	59	57	−2
Waterhead‡	NX 752 836	200	33	36	+3
Urban site					
Glasgow	NS 608 687	100	55	39	−16

* From equation (1).
† 1979–80 only.
‡ 1980 only.

to the ionic composition of rainfall collected in our gauges has been independently estimated, and is discussed later.

Measurements of *p*H and the concentrations of sulphate and magnesium ions were made on samples collected during 1978 and 1979; since 1980 the range of analyses has included concentrations of nitrate, chloride, sodium, potassium and ammonium ions (J.N.C. and D.F., work in preparation). *p*H was measured taking the necessary precautions for samples of low ionic strength[12] using standard glass and calomel electrodes. Chloride concentrations were measured using an ion-selective electrode. Sulphate analysis was by the continuous-flow thoron method[13] with aqueous isopropanol as solvent. Measurement of nitrate was by reduction to nitrite followed by complexation with sulphanilamide/*N*-1-naphthylethylenediamine dihydrochloride[14] in a continuous-flow system, and ammonium was measured as ammonia using continuous flow through an ammonia-selective electrode. Magnesium and calcium concentrations were determined by atomic absorption spectroscopy (after addition of lanthanum) and sodium and potassium concentrations by flame emission spectroscopy.

Fig. 2 Annual deposition of hydrogen ions (kg H^+ ha^{-1}) in rain over northern Britain.

Magnesium ion concentrations were used to estimate the sea-salt contribution to total ionic composition. Fifteen of the 16 rain-collecting sites were located in rural areas remote from local sources of air pollutants, with one site located in central Glasgow (Fig. 1). They provided an approximately uniform coverage of northern Britain (Table 1).

The surfaces of rain collectors are contaminated by dry deposition of gaseous SO_2 and NO_2 and particulate sulphate. As a result the deposition of sulphur and hydrogen ions in rain is overestimated, their estimates in southeastern Scotland being exaggerated by ~10% and in northern and western Scotland by 2–5%. The contamination was quantified from funnel washings following dry days and from laboratory experiments in which funnels were washed after exposure to SO_2 (unpublished data). Collectors sited at 2 m above ground, to limit contamination by surface-derived material, capture 5–10% less rain than identical collectors at ground level[15]. However, for estimates of annual deposition this error can be minimized by combining the solute concentrations found in the network with the amounts of rain recorded by the Meteorological Office.

The weighted mean acidity of rain was least in the north-west of northern Britain (20 μEquiv. H^+ l^{-1}, *p*H 4.7) and greatest in the east and south of northern Britain (60 μEquiv. H^+ l^{-1}, *p*H 4.2) (Fig. 1) The regional gradient, which is relatively simple, is best described by the following quadratic relationship between National Grid coordinates and hydrogen ion concentrations:

$$[H^+]\,(\mu\text{Equiv. } l^{-1}) = 3.80\times10^{-6}\,E^2 + 7.05\times10^{-7}\,E.N. - 3.16\times10^{-6}\,N^2 - 1.06\times10^{-2}\,E + 4.32\times10^{-2}\,N - 118 \quad (1)$$

where *N* and *E* are four-figure northings and eastings from Ordnance Survey maps.

This empirical equation accounts for 94% of the variation,

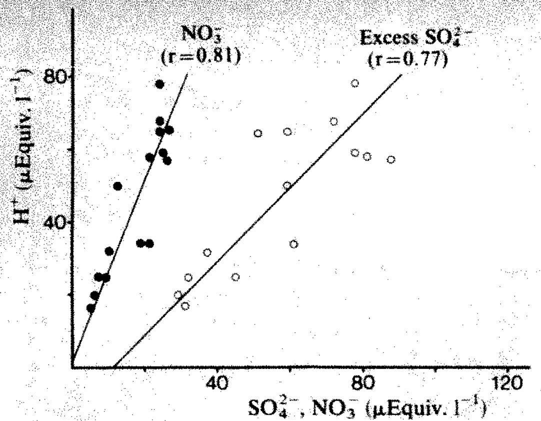


Fig. 3 Relationship between hydrogen ion concentrations in rain collected at 15 rural sites in northern Britain for the period 1978–80, and concentrations of nitrate (●) and non-marine (excess) sulphate (○) ions.

the average uncertainty of the predicted values for the locations marked in Fig. 1 being ±9% (Table 1).

As the regional pattern has been established from collections made in rural locations, it is of interest to consider local variations associated with major industrial and residential areas. To provide an estimate of the 'city influence' the Glasgow site at which the annual average SO_2 concentration is ~80 μg m^{-3} (ref. 16) was included in the study. The regional pattern predicts a rainfall acidity in Glasgow equivalent to 39 μEquiv. $H^+ l^{-1}$ (pH 4.45) whereas an acidity of 55 μEquiv. $H^+ l^{-1}$ (pH 4.26) was actually recorded, indicating that ~30% of the acidity at that site was of local origin. The difference between the predicted regional value and measured local value exceeds the largest difference between predicted and measured values for the rural network sites by a factor of 2 (Table 1). The excess sulphate concentration (that sulphate remaining when sea-derived sulphate has been subtracted from the total) is also large in Glasgow, being approximately twice the regional pattern estimate (92 compared with 43 μEquiv. l^{-1}). This suggests that some of the locally derived sulphate does not contribute to acidity, but seems to be associated with ammonium ions, for which the Glasgow concentration was over twice the average for northern Britain (43 compared with 19 μEquiv. l^{-1}).

When the hydrogen ion concentrations for rain falling at the centre of each 10 km square of the National Grid are combined with the Meteorological Office's 30 yr data for amounts of precipitation, it can be seen that the deposition of hydrogen ions is strongly influenced by height above sea level (Fig. 2), an effect mainly attributable to the relation between amounts of rain and altitude over northern Britain. Rainfall during the period of study exceeded the 30-yr average by ~10%[17]. Rainfall chemistry is not independent of altitude, fine rain often encountered at high altitudes having characteristically larger ionic concentrations[18]. Single rain events show a decrease in concentration with the amount of rain[19–21]. As the combined effect of these two processes is uncertain, average rainfall chemistry has been assumed to be constant with altitude. It is clear from Fig. 2 that rainfall in the west more than compensates for smaller average acidities, so that although most of the region has a hydrogen ion input in rain of ~0.5 kg H^+ ha^{-1} yr^{-1}, there are areas with much larger inputs where annual hydrogen ion deposition may exceed 1 kg ha^{-1} yr^{-1}. For a consideration of effects the degree of acidity is important, but the quantity of acid deposited may be equally important. Henriksen[22] suggested that the effects of the direct inputs of acidity in rain, when acidifying bodies of freshwater, could be likened to a large-scale titration. The deposition pattern shown in Fig. 2 identifies areas of northern Britain where inputs of acidity are large and comparable with areas of Norway where fish populations have decreased markedly[3].

The dominant anions in rain are sulphate, nitrate and chloride. Seawater is a significant source of chloride and sulphate ions in rain, but makes a negligible contribution to nitrate or hydrogen ion concentrations. Our evidence suggests that most chloride has a marine origin, and therefore does not contribute to acidity, although non-marine chloride has been detected at a site in central Scotland close to known sources of gaseous HCl (unpublished data). The proportion of marine sulphate varies from >60% at coastal western sites to 10% in the east. The two major anions associated with acidity are therefore nitrate and excess sulphate, each of which is strongly correlated with hydrogen ion (Fig. 3). Similar correlations have been noted by others[23,24]. There are two extreme interpretations of these correlations. On the one hand, all of the nitrate could be associated with hydrogen ions, in which case 38% of the acidity would be nitric acid, the remaining 62% being sulphuric acid. At the other extreme there is sufficient excess sulphate to account for all of the measured acidity as sulphuric acid. However, because ammonium sulphate is frequently the major component of sub-micrometre aerosols in the United Kingdom[25] and because sulphate concentrations in rain are closely associated with aerosol sulphate concentrations[26], a significant proportion of the excess sulphate in rain is likely to occur as ammonium sulphate. This suggests that at least some of the acidity is associated with nitrate ions and that the second interpretation is unlikely to be tenable. On average an increase in hydrogen ion concentration of 10 μEquiv. l^{-1} was associated with increases in nitrate and sulphate concentrations of 4 and 10 μEquiv. l^{-1} respectively, which suggests an average contribution to acidity of 29% by nitric acid and 71% by sulphuric acid. This is consistent with conclusions reached for the USA[27] and Scandinavia[28].

The acidity of rain falling in the east and south-east of northern Britain is comparable with that in areas of Scandinavia and North America where populations of fish have been severely depleted[3,4]. The regional pattern of acidity shows no correlation with population density nor with the distribution of major sources of sulphur dioxide and nitrogen oxides within the region[29]. Associated studies have shown the acidity to be highly episodic with a complex frequency distribution, and that the most acid episodes (pH<4.0) arise from air masses that have had a trajectory over a major industrial region of the UK or Europe (D.F. and J.N.C., work in preparation). This implies that the scales of transport of the substances which acidify rain are much larger than the scale of northern Britain, and that sources within this region only influence the regional distribution of acidity in rain to a minor extent.

We thank the Forestry Commission, Nature Conservancy Council, Department of Agriculture and Fisheries for Scotland, Hill Farming Research Organisation, Freshwater Biological Association, Glasgow University, Ministry of Agriculture, Fisheries and Food, and Glasgow District Council for assistance with the collection of rain samples, Mr J. N. R. Jeffers and Professor F. T. Last for their helpful comments, and Miss M. McPherson for technical assistance. This work was financed by the NERC, the EEC and the UK Department of the Environment.

Received 16 December 1981; accepted 26 March 1982.

1. Sweden's case to the *U.N. Conf. on the Human Environment* (Royal Ministry of Foreign Affairs, Stockholm, 1971).
2. Leivestad, H., Hendrey, G., Muniz, I. P. & Snekvik, E. in *SNSF Report FR6/76* (ed. Braekke, F. H.) 87–111 (SNSF, Oslo 1976).
3. Muniz, I. P. & Leivestad, H. in *Ecological Imapct of Acid Precipitation* (eds Drabløs, D. & Tollan, A.) 84–92 (SNSF, Oslo, 1980).
4. Harvey, H. H. in *Ecological Impact of Acid Precipitation*, (eds Drabløs, D. & Tollan, A.) 93–98 (SNSF, Oslo, 1980).
5. Harriman, R. & Morrison, B. in *Ecological Impact of Acid Precipitation* (eds Drabløs, D. & Tollan, A.) 312–313 (SNSF, Oslo, 1980).
6. Lawes, J. B., Gilbert, J. H. & Pugh, E. *Phil. Trans. R. Soc.* **II**, 431–577 (1861).
7. Gorham, E. *Phil. Trans. R. Soc.* **B241**, 147–178 (1958).
8. Stevenson, C. M. *Q. J R. met. Soc.* **96**, 56–70 (1968).
9. Martin, A. & Barber, F. R. *Atmos. Envir.* **12**, 1481–1487 (1978).
10. Martin, A. *Sci. Total Envir.* **13**, 119–130 (1979).
11. Galloway, J. N. & Likens, G. E. in *Proc. 1st int. Symp. Acid Precipitation and Forest Ecosystem* (eds Dochinger, L. S. & Seliga, T. A.) 137–155 (USDA Forest Service General Technical Report NE-23, 1976).
12. Galloway, J. N., Crosby, B. J. & Likens, G. E. *Limnol. Oceanogr.* **24**, 1161–1165 (1979).
13. Persson, G. A. *Air Water Pollut. int. J.* **10**, 845–852 (1966).

14. Eggleton, A. E. J. & Atkins, D. H. *Results of the Tees-side Investigation*, 165 (AERE-R 6983, HMSO, London 1972).
15. Robinson, A. C. & Rodda, J. C. *Met. Mag.* **98**, 113–120 (1969).
16. *National Survey Smoke and Sulphur Dioxide* (Warren Spring Laboratory, Department of Industry, 1979).
17. *Isopercental Rainfall Maps for U.K. 1978–1980* (UK Meteorological Office, to be published).
18. Tomlinson, G. H., Brouzes, R. J. P., McLean, R. A. N. & Kadlecek, J. in *Ecological Impact of Acid Precipitation* (eds Drabløs, D. & Tollan, A.) 134–137 (SNSF, Oslo, 1980).
19. Granat, L. *Atmos. Envir.* **12**, 413–424 (1978).
20. Hicks, B. B. & Shannon, J. D. *J. Am. met. Soc.* **18**, 1415–1420 (1979).
21. Fowler, D. in *Ecological Impact of Acid Precipitation* (eds Drabløs, D. & Tollan, A.) 22–32 (SNSF, Oslo, 1980).
22. Henriksen, A. in *Ecological Impact of Acid Precipitation* (eds Drabløs, D. & Tollan, A.) 68–74 (SNSF, Oslo, 1980).
23. Marsh, A. R. W. *Atmos. Envir.* **12**, 401–406 (1978).
24. Glover, G. M., Kallend, A. S., Marsh, A. R. W. & Webb, A. H. in *Effects of Acid Precipitation on Terrestrial Ecosystems* (eds Hutchinson, T. C. & Havas, M.) 95–110 (Plenum, New York, 1980).
25. Heard, M. J. & Wiffen, R. D. *Atmos. Envir.* **3**, 337–340 (1969).
26. Garland, J. A. *Atmos. Envir.* **12**, 349–362 (1978).
27. Galloway, J. N. & Likens, G. E. *Atmos. Envir.* **15**, 1081–1085 (1981).
28. Overrein, L. N., Seip, H. M. & Tollan, A. (eds) *Final Rep. SNSF Project, 1972–1980*, 23–32 (SNSF, Oslo, 1980).
29. Fowler, D. & Cape, J. N. in *Effects of Gaseous Air Pollutants in Agriculture and Horticulture* (eds Unsworth, M. H. & Ormrod, D. P.) (Butterworths, London, in the press).

Geomagnetic secular variation as a precursor of climatic change

V. Courtillot, J. L. Le Mouel & J. Ducruix

Département des Sciences de la Terre, Université Paris VII and Institut de Physique du Globe, Université Paris VI, 2 Place Jussieu, 75230 Paris Cedex 05, France

A. Cazenave

Groupe de Recherches de Géodésie Spatiale, CNES, Toulouse, France

Long period trends in climate are usually associated with solar disturbance[1]. For example, a close similarity has been demonstrated[2,3] between variations in the length of day (LOD) with periods greater than about 10 yr and trends in several climatic indices over the past 150 yr. We point out here a correlation between variations in the Earth's magnetic field, the Earth's rotation rate and some climatic indicators, thus suggesting a possible long term influence of core motions on climate. We suggest that geomagnetic secular variation can be used to forecast a climatic change in the 1990s.

A good correlation has been established[2,3] between LOD variations and either global temperature θ (correlation coefficient 0.85; θ lagging LOD variations by ~5 yr, see Fig. 1) or the long period atmospheric and oceanic excitation function ψ^T related to near surface winds, changes in the atmospheric mass distribution and eustatic changes in sea level (correlation coefficient 0.75; ψ^T lagging LOD variations by ~15 yr see Fig. 9.6 in ref. 4; ψ^T is computed from ground level

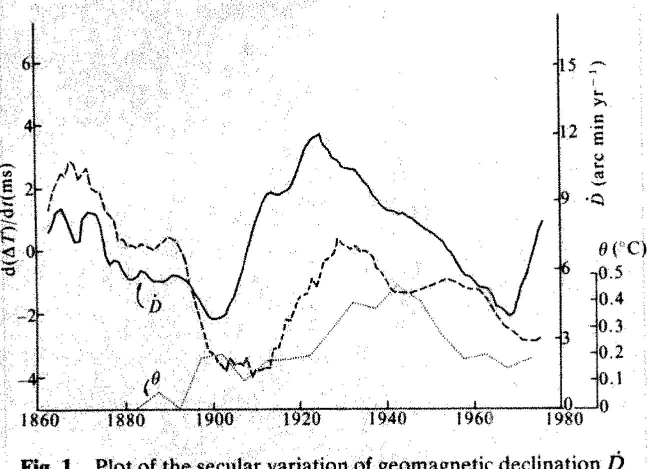


Fig. 1 Plot of the secular variation of geomagnetic declination $\dot{D}$ (solid curve), of the excess length of day $d(\Delta T)/dt$ (dashed curve) and of successive 5 yr averages of the temperature θ over the whole Earth (expressed as departures from the means for 1880–84, after Fig. 18.2 of ref. 1, dotted curve).

pressure sea-level data under the geostrophic assumption). Periods of increasing zonal circulation and increasing global surface temperature are found to correspond to periods of Earth acceleration, while periods of decreasing zonal circulation and decreasing global surface temperature correspond to periods of Earth deceleration. It has been further observed[3] that the excitation function ψ^T represents only about 10% of the amount necessary to explain the LOD variations. This led to the conclusion that it is not possible as yet to establish causal relationships between these phenomena and that one cannot decide whether (1) atmospheric circulation causes the long period changes in LOD, or (2) the fluctuations are both a consequence of a third phenomenon, or (3) the fluctuations in LOD cause the observed variations in the atmospheric circulation.

Although A.C. and Lambeck[3] note the problems related to phase lag between θ (or ψ^T) and LOD and the insufficient amplitude of the excitation function, they tend to reject hypothesis (3) because "the total mechanism for exciting the LOD changes remains to be explained".

More recently, Lambeck[4] has reviewed the long period or decade fluctuations in the Earth's rotation and discussed their geophysical origin. He notes that the role of the core is central, for it is the only sufficiently mobile part of the Earth with sufficient mass to modify the rotation by the observed amounts on that time scale. Out of the mechanisms proposed for core–mantle coupling, pressure or inertial coupling due to the ellipticity of the core boundary, topographic coupling due to bumps on the core–mantle interface and viscous coupling all seem to be inadequate[4,5]. Only electromagnetic core–mantle coupling, as suggested by Bullard *et al.*[6], survives the screening process despite severe uncertainties regarding the conductivity of the lower mantle and the core motions. The key observations required to resolve these uncertainties come from observations of the secular variation, in particular the rate of westward drift, of the geomagnetic field. The magnetic observations available to Lambeck did suggest some kind of correlation between LOD and geomagnetic variations, but the evidence was inconclusive.

The situation has been improved with the observation[7] that the secular variation of the geomagnetic field had undergone sharp accelerations of global extent. The last and best studied such phenomenon which occurred around 1970 is described in refs 8, 9, 14. The secular acceleration can first be used to infer an upper bound on the electrical conductivity of the lower mantle which is found to be of the order of a few hundred $\Omega^{-1}\,m^{-1}$ (refs 10, 11). This parameter is essential in any computation related to electromagnetic core–mantle coupling.

The correlation[7] between secular variation accelerations and extrema in LOD fluctuations has recently been put on a firmer and more systematic basis[12] with the demonstration of a good correlation between LOD variations and the secular variation of declination, observed for instance in European observatories where the longest records are available: the correlation for the period 1865–1975 is far more convincing than any previously reported (Fig. 1). The correlation coefficient is found to be about 0.8, with the magnetic variations leading LOD fluctuations by about 10 yr. J.L. Le M. and V.C.[13] suggest that a model of accelerated westward drift of the upper core based on the original model of Bullard *et al.*[6] accounts both qualitatively and quantitatively for the observations. In this model, the Earth's mantle, an upper core layer and the lower core are all electromagnetically coupled. A sudden torque applied at the time of a secular variation acceleration on a ~100 km thick upper core layer leads to a very fast response of westward drift as observed at the surface, and to a fast response of the lower core in the opposite sense. The lower core finally drags the mantle with a longer time-constant, of the order of 10 yr, as is typically observed in the decade fluctuations of LOD. Thus, these decade fluctuations are probably caused by rearrangements in core motions. Having found a mechanism which can excite LOD changes, it becomes natural to accept the hypothesis that fluctuations in LOD in turn are responsible for changes in atmospheric circulation (although the hypothesis of an indepen-

0028-0836/82/220386—02$01.00

dant common cause cannot be rejected). The observed phase lag between ψ^T and LOD and the amplitude of ψ^T are then at least qualitatively accounted for. There is an attractive causal chain whereby, over the decade period range, angular momentum is transferred from the core to the mantle and finally to the atmosphere (although details of the transfer mechanism remain to be studied).

Lambeck and A.C.[3] point out that LOD variations can be used as an indicator of future climatic trends. Lamb (ref. 1, p. 710) in his review of scientifically based climate forecasts notes that variations are commonly associated with long term variations of solar disturbance. He notes a considerable measure of agreement between the various forecasts, with a trend towards colder climates with weakened general atmospheric circulation which is expected to continue into the twenty-first century, in some cases with a sharp further cooling about 1980 and easier conditions for a time in the first half of the next century. The geomagnetic results referred to here suggest that changes in secular variation can be used as indicators of future LOD fluctuations[12,13] and, with an even longer time lead of 15–25 yr, as indicators of climatic changes, in particular changes in global surface temperature. The clearly established 1970 secular acceleration, which has now been maintained for a decade, suggests a correlated positive acceleration in LOD in the immediate future and the beginning of a period of steady increase in average global temperature around 1990 (±5 yr). K. Brian (personal communication) has recently pointed out a possible correlation between the frequency of geomagnetic reversals and deep ocean temperature over the past 55 Myr, thus suggesting another (very) long term relationship between climate and the Earth's magnetic field.

This work is IPG contribution no. 582.

Received 9 February; accepted 21 April 1982.

1. Lamb, H. H. *Climate, Present, Past and Future*, Vol. 2 (Methuen, London, 1977).
2. Cazenave, A. thesis, Univ. Toulouse (1975).
3. Lambeck, K. & Cazenave, A. *Geophys. J. R. astr. Soc.* **46**, 555–573 (1976).
4. Lambeck, *The Earth's Variable Rotation. Geophysical Causes and Consequences* (Cambridge University Press, London, 1980).
5. Rochester, M. G. *IAGA Bull.* **43**, 115 (1979).
6. Bullard, E. C., Freedman, C., Gellman, H. & Nixon, J. *Phil. Trans. R. Soc.* A **243**, 67–92 (1950).
7. Courtillot, V., Ducruix, J. & Le Mouël, J. L. *C.R. hebd. Séanc. Acad. Sci., Paris* D **287**, 1095–1098 (1978); B **289**, 173–175 (1979).
8. Ducruix, J., Courtillot, V. & Le Mouël, J. L. *Geophys. J. Roy. astr. Soc.* **61**, 73–94 (1980).
9. Ha Duyen, C., Ducruix, J. & Le Mouël, J. L. *C.R. hebd. Séanc. Acad. Sci., Paris* B **293**, 157–160 (1980).
10. Achache, J., Courtillot, V., Ducruix, J. & Le Mouël, J. L. *Phys. Earth planet. Inter.* **23**, 72–75 (1980).
11. Achache, J., Le Mouël, J. L. & Courtillot, V. *Geophys. J. R. astr. Soc.* **65**, 579–601 (1981).
12. Le Mouël, J. L., Madden, T. R., Ducruix, J. & Courtillot, V. *Nature* **290**, 763–765 (1981).
13. Le Mouël, J. L. & Courtillot, V. *Phys. Earth planet. Inter.* **24**, 236–241 (1981); *J. geophys. Res.* (in the press).
14. Malin, S. R. C. & Hodder, B. M. *Nature* **296**, 726–728 (1982).

Spore–pollen evidence for early Oligocene high-latitude cool climatic episode in northern Canada

Geoffrey Norris

Department of Geology, University of Toronto, Toronto, Ontario, Canada M5S 1A1

Climatic cooling in the Eocene leading to markedly lower temperatures in the Oligocene has been documented in oxygen isotope studies in the North Pacific[1,2], New Zealand[3], South Pacific[4] and North Sea[5]. These lower temperatures in the Oligocene correspond to palaeoclimatic interpretations of North American leaf floras[6,7] which suggest a profound cooling occurring towards the end of the Eocene, indicated also in the southeastern United States from studies on mid-Tertiary spores and pollen[8]. Recently, Collinson *et al.*[9] have shown palaeobotanically that in southern England cooling occurred gradually starting in the latest early Eocene leading to two major periods of floristic change before the end of the Eocene. Thick Tertiary sections in the subsurface Mackenzie Delta region, Northwest Territories[10,11] provide well-preserved Palaeogene palynofloras[12] discussed here which indicate a cooler climate in the Oligocene following warm-temperate conditions in the Eocene. This widespread early–middle Oligocene cool episode is thus represented in both high-latitude and lower-latitude floras and seems to have been of global extent, persisting until towards the end of the Oligocene, but its precise dating remains a problem. In northern Canada it is followed by an amelioration of the climate in the late Oligocene which persisted probably until the middle Miocene. The rate of cooling during the Eocene and Oligocene may have varied depending on locality, but once established the cooler Oligocene climate exerted a profound influence on biotic development.

The subsurface Mackenzie Delta region (lat. 69 °N, long. 134 °W) includes a thick Tertiary molasse pile[11,13,14] deposited in the Richards Island Basin, representing several regressive depositional events formed by prograding deltas with marine deposits at the base. A relatively continuous Tertiary section is available here from the Palaeocene through to the Quaternary[11,12,15,16] but a major unconformity removes much of the Palaeogene in the southern part of the basin. The Imperial Nuktak C-22 well in the northern part of the basin has a relatively complete Tertiary section (Fig. 1) down to TD 12,650 ft in the middle Eocene part of the Richards Formation[12], a prodeltaic mudstone unit in excess of 5,000 ft thick which passes laterally southwards into the upper part of the Palaeocene–middle Eocene deltaic Reindeer Formation[11]. The Richards Formation is largely middle to late Eocene on the basis of the *Haplophragmoides* foraminiferal assemblage[11] and the presence near the base[12] of *Pistillipollenites mcgregorii* Rouse and the dinoflagellate cysts *Glaphyrocysta ordinata* (Will., and Dow) Stover and Evitt, *Cordosphaeridium gracile* (Eis.) Dav. and Will., and *Wetzeliella* sp. cf. *W. hampdenensis* Wilson. The precise placement of the Eocene–Oligocene boundary has not been achieved yet because ranges of terrestrial palynomorphs are uncertain at these high latitudes[12,15,17]. It probably occurs almost 1,700 ft below the top of the Richards Formation (Fig. 1) if Rouse[17] is correct in limiting the range of *Integricorpus* sp. (which probably does not occur below 9,500 ft in the Nuktak C-22 well) to the Oligocene. This interpretation is supported by the presence in the upper 1,700 ft of the Richards Formation of spore–pollen species from the putative Lower Oligocene of northeastern China[18].

The terrestrial flora of the middle–late Eocene part of the Richards Formation is diverse and contains taxa which can be

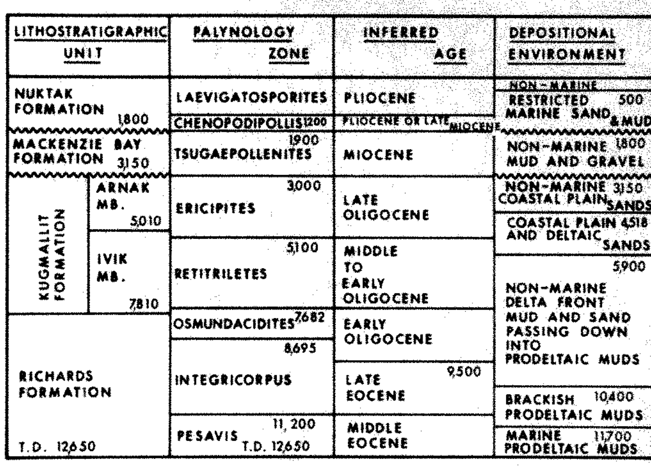


Fig. 1 Tertiary stratigraphy of Imperial Nuktak C-22 well. Palynostratigraphy, zonal names and age according to ref. 12, but Eocene–Oligocene boundary may occur lower (see text for discussion). Lithostratigraphy after Young and McNeil[11]. Numbers give the depth in feet.

0028–0836/82/220387—03$01.00

attributed to the following extant genera as discussed in detail elsewhere[8,12,15,19,20]: *Pinus, Sequoia, Metasequoia, Ulmus, Tilia, Castanea, Quercus, Tsuga, Pterocarya, Picea, Alnus, Betula* and *Sphagnum*. The latter four genera range into cool temperate or boreal regions at the present, but the remaining nine taxa are clearly temperate or warm temperate. This mixed palynoflora may be due to long-distance pollen transport from cooler—perhaps mountainous—regions of the northern Cordillera into warm temperate lowlands prevailing in the Mackenzie Delta region, a possibility supported by the sedimentology[13]. Mixing due to cavings in the cuttings samples may also introduce non-indigenous species into the samples, indicated by the lack of *Picea* and *Betula* in core samples compared with cuttings samples of the Richards Formation. Thus the indigenous palynoflora of the Richards Formation indicates temperate or warm temperate climates in the middle and late Eocene, possibly persisting into the earliest Oligocene. A sharp decrease in diversity of palynofloras and extinction of several species occurs within the prodeltaic mudstones of the early Oligocene part of the Richards Formation between 9,600 and 8,900 ft and is not associated with a major facies change. The Ivik Member of the superadjacent Kugmallit Formation contains a less diverse palynoflora of notably different composition, particularly in the upper part of the member, compared with the Richards Formation. Miospores attributed to extant taxa in the Ivik floras include *Sphagnum, Picea, Pinus, Lycopodium, Osmunda, Tsuga, Alnus, Betula, Carpinus* and *Pterocarya*. However, it is significant that a variety of fungal spores and hyphae, bryophytes and pteridophytes, *Sequoia, Ulmus, Tilia, Quercus, Castanea*, and other thermophylic angiosperms which are common in the Eocene–earliest Oligocene Richards Formation disappear in the lower half of the early–middle Oligocene[12] Ivik Member (below 6,500 ft) but reappear higher in the late Oligocene[11,12] Arnak Member of the Kugmallit Formation accompanied by ericaceous pollen amongst others. The lower Ivik palynofloral changes occur in an interval of rhythmic alterations of mudstone and sandstone in medium-scale coarsening-upward trends typical of progradational deltas. No major facies change occurs in this interval. The lack of thermophylic angiosperm taxa in part of the early and perhaps middle Oligocene indicates a significant cool interval, possibly cooler than the climate of the present Great Lakes region where these taxa terminate their northern ranges in the pollen rain[21]. Their resurgence in the Arnak Member is accompanied by an overall average doubling of terrestrial species numbers. This cannot be attributed entirely to a change to alluvial or deltaic conditions from the delta-front conditions of the Ivik Member because other similar progradational cycles in the Richards Island Basin[11] are not characterized by such a change in palynofloral diversity (unpublished data). The increased diversity, however, may be related to an amelioration of climate as warm temperate conditions were re-established.

The reasons for this early to middle Oligocene cool interval are not clear. An explanation in terms of a change in obliquity of the ecliptic[6,7] is not considered dynamically acceptable[22], nor is it in accord with other floral and faunal evidence of pronounced Palaeogene seasonality indicating significant annual periods of winter darkness at high latitude[23]. However, a cooler interval during the Oligocene appears to be of global extent, affecting oceanic water temperatures in both hemispheres[1-4] affecting the marine biota[22,24], and affecting high and low latitude terrestrial floras for at least 5 Myr. In Central Europe, however, the coolest mid-Tertiary terrestrial floras are late Oligocene[25], in contrast to the early and middle Oligocene cooling indicated in many regions as outlined above. In central British Columbia, Piel[26] reported humid warm temperate to sub-tropical climates for the putative early Oligocene Australian Creek Formation of central British Columbia based on the presence of miospores attributed to *Metasequoia, Picea, Pinus, Podocarpus, Acer, Betula, Carya, Castanea, Engelhardtia, Fagus, Fraxinus, Juglans, Liquidambar, Nyssa, Tilia* and *Ulmus/Zelkova*. Rouse and Mathews[27], however, have interpreted this assemblage to be late early Oligocene representing a mixed mesophytic warm temperate forest, similar to those of southeastern United States or the upper Yangtze River Valley of China. The age of these palynofloras is contentious but may be indicated by the report of associated titanothere teeth (a brontotherine genus, possibly *Megacerops* or *Brontotherium*) said to be late early Oligocene[27]. On the other hand, the presence here of species of *Parviprojectus, Pesavis, Ctenosporites* and *Striadiporites* points to a correlation with the putatively middle and late Eocene *Pesavis* and *Integricorpus* zones in the Richards Island Basin[12] (Fig. 1). Climatic gradient may have strongly affected the differential attenuation of species ranges at different latitudes in the late Palaeocene, thus allowing the extension of arctic Eocene palynomorph species into the early Oligocene at lower latitudes. Until this matter is resolved, however, the possibility exists that the *Integricorpus* zone may be entirely early Oligocene and, therefore, that the climatic cooling indicated by palynofloral data in the upper part of the *Integricorpus* zone and higher in the western Arctic may have occurred later than the Eocene–Oligocene boundary.

This dramatic temperature drop in the Oligocene leads to a paradox[28]. Cooler global air and water temperatures would lead to decreased evaporation rates on a global scale, but, nevertheless, moisture continued to be transported to Antarctica where ice continued to accumulate and reached sea level at the end of the Oligocene. Ice accumulation in the Arctic Ocean did not commence until some time in the Pliocene[29] when the Miocene temperate terrestrial floras were replaced by boreal tundra floras[12], and at which time a major episode of molluscan extinction occurred globally, probably in response to a lethal decline in temperature[30].

The palaeooceanographical changes at the Eocene–Oligocene boundary have been related[24] to cooler air temperatures and floral changes on land. O'Keefe's hypothesis[22] for early Oligocene cooling postulates the development of tektite ring system around the Earth which cast a shadow on the winter hemisphere. This would lead to a temperature drop of ~20 °C and remain stable for several million years. Synchronous global cooling could also result from a variation in the solar output[23]. Progress in understanding the mechanisms will depend in part on increased precision in dating Palaeogene terrestrial deposits. The exact timing of terrestrial floral changes from region to region must be determined and compared with biotic and isotopic changes in the oceans. If the floral and faunal changes are indeed synchronous globally, a pervasive and almost instantaneous mechanism must be sought. If the terrestrial floral changes accompanying cooling prove to be diachronous, a mechanism dependent on local or regional changes in geography, oceanic circulation, or tectonics may be implicated.

This work was supported by an operating grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada, a contract from the Department of Supply and Services, Ottawa and a NSERC strategic energy grant for Basin Analysis at the University of Toronto. Samples and data were provided by the Institute of Sedimentary and Petroleum Geology, Calgary, Austin and Cumming Exploration Consultants, and Imperial Oil Ltd. Dr Glenn Rouse provided valuable information and discussion.

Received 14 January; accepted 5 April 1982.

1. Douglas, R. G. & Savin, S. M. *Init. Rep. DSDP Leg.* **32,** 509–520 (1975).
2. Savin, R. J. *A. Rev. Earth planet. Sci.* **5,** 319–355 (1977).
3. Devereux, I. *New Z. J. Sci.* **10,** 988–1101 (1967).
4. Shackleton, N. J. & Kennett, J. P. *Init. Rep. DSDP Leg.* **29,** 743–755 (1975).
5. Buchardt, B. *Nature,* **275,** 121–123 (1978); *Dansk. Met. Inst. Climatol. Pap.* 4, 243–248 (1978).
6. Wolfe, J. A. *Am. Scient.* **66,** 694–703 (1978).
7. Wolfe, J. A. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **30,** 313–323 (1980).
8. Frederiksen, N. O. *J. Paleont.* **54,** 728–739 (1980).
9. Collinson, M. E., Fowler, K. & Boulter, M. C. *Nature* **291,** 315–317 (1981).
10. Young, F. G. Myhr, D. W. & Yorath, C. J. *Geol. Surv. Can. Pap.* 76–11 1–65 (1976).
11. Young, F. G. & McNeil, D. W. *Geol. Surv. Can. Bull.* (in the press).
12. Norris, G. *Geol. Surv. Can. Bull.* **340** (in the press).
13. Dixon, J. *Geol. Surv. Can. Pap.* **80–81,** 1–11 (1981).
14. Bowerman, J. N. & Coffman, R. C. *Can. Soc. petrol. Geol. Mem.* **4,** 649–662 (1975).

15. Staplin, F. L. *Bull. Can. petrol. Geol.* **24,** 117–36 (1976).
16. Ioannides, N. S. & McIntyre, D. J. *Geol. Surv. Can. Pap.* **80–1A,** 197–208, (1980).
17. Rouse, G. E. *Am. Ass. Strat. Palynol., Contr. Ser.* **5A,** 48–65 (1977).
18. Sung, Z. C. *et al. Nanjing Inst. Geol. Paleont., Chinese Acad. Sci* 1–177 (Kexue Chubanshe, Peking, 1978).
19. Gruas, Cavagnetto, C. *Mem. Soc. Geol. Fr.* **131,** 1–64 (1978).
20. Elsik, W. C. *Palaeontographica, Abt. B.* **149,** 90–111 (1974).
21. Webb, I. & McAndrews, J. H. *Mem. geol. Soc. Am.* **145,** 267–299 (1976).
22. O'Keefe, J. A. *Nature* **285,** 309–311 (1980).
23. McKenna, M. C. *Palaeogeogr., Palaeoclimatol. Palaeoecol.* **30,** 349–362 (1980).
24. Keigwin, L. D. *Nature* **287,** 722–725 (1980).
25. Hochuli, P. *Ann. Geol. Pays Hellen., Tome hors ser. fasc.* **II,** 515–523 (1979); *Beitr. Palaeont. Osterr.* **4,** 1–132 (1978).
26. Piel, K. M. *Can. J. Bot.* **49,** 1885–1920 (1971).
27. Rouse, G. E. & Mathews, W. H. *Bull. Can. petrol. Geol.* **27,** 418–445 (1979).
28. Frakes, L. A. *Climates Throughout Geologic Time,* 1–310 (Elsevier, Amsterdam, 1979).
29. Clarke, D. L. *Bull. geol. Soc. Am.* **82,** 3313–3324 (1971).
30. Stanley, S. M. & Campbell, L. D. *Nature* **293,** 457–459 (1981).

Carbon cycle changes of the Zechstein Sea: isotopic transition zone in the Marl Slate

M. Magaritz

Isotope Department, The Weizmann Institute of Science, Rehovot, Israel

P. Turner

Department of Geological Sciences, The University of Aston, Gosta Green, Birmingham B4 7ET, UK

The carbon isotope composition of carbonate rocks is one parameter used to study the carbon, sulphur and oxygen cycles through geological time[1,2]. The understanding of these interrelated cycles is important in quantitative flux modelling of sedimentary rocks. The carbon isotope composition in the large reservoir of dissolved carbonate in ocean water is a result of mass balance between the amount of carbon in the reduced organic reservoir and the oxidized carbonate reservoir at any given stage. The reduced carbon reservoir preferentially sequesters ^{12}C owing to biological fractionation processes. Any change in the removal ratio of oxidized/reduced carbon will be reflected by the change of the δ^{13}C value of the oceanic carbonate system. Here we present a detailed study of the change in ^{13}C which took place at the base of the Upper Permian Zechstein sequence.

On the basis of measurements of carbon isotopes on rocks of various ages, early models assumed that the mean δ^{13}C (carbonate) values were constant[3–6]. In recent years, however, descriptions of secular variation of δ^{13}C values in carbonate rocks during the Mesozoic[7,8] and Cenozoic[9,10] have become established. These secular variations are explained by large scale burial of organic material and are associated with ocean wide changes such as formation of anoxic conditions and sea-level fluctuation[11].

The same mechanism is used in modelling mass balance chemical and isotopic variation in the C–S–O system between geological periods[1–3]. These mass balance models[2,3] assume that high δ^{13}C values prevail for the whole of the Permian period[12] but, in fact, the detailed evidence indicates that only certain parts of the Permian are characterized by these high values[13]. The ^{13}C maximum in carbonate rocks discussed here starts at the base of the Upper Permian Zechstein sequence and lasts up to the top of the third Zechstein cycle[14] when fully marine conditions terminated. Because such a change also means changes in the amount of oxygen tied to the carbon system then it may be expected to be reflected in the oxygen cycle and the cycles of other elements to which oxygen is related (sulphur and iron).

Some of the carbonate rocks of the Permian period exhibit anomalous enrichment in ^{13}C relative to other Phanerozoic rocks. These high δ^{13}C values are observed even if one mixes all the Permian rocks analysed and compared them with rocks from other periods[12,15]. The fact that some of the Permian carbonate rocks from various parts of the world reach values of up to +7‰ (relative to the normal fluctuation range −1 to +3‰)[16–19] indicates that extreme changes in the C–O–S cycle took place. These changes had also been predicted from the anomalously low values of δ^{34}S in sulphates[1,20]. These two values (δ^{13}C and δ^{34}S) are negatively correlated if one uses mean values for periods. The model showing the relation between the carbon and sulphur cycle based on isotopic data, assuming scales for mixing in the sedimentary cycle of the order of 120–240 Myr, is given by Schidlowski *et al.*[3].

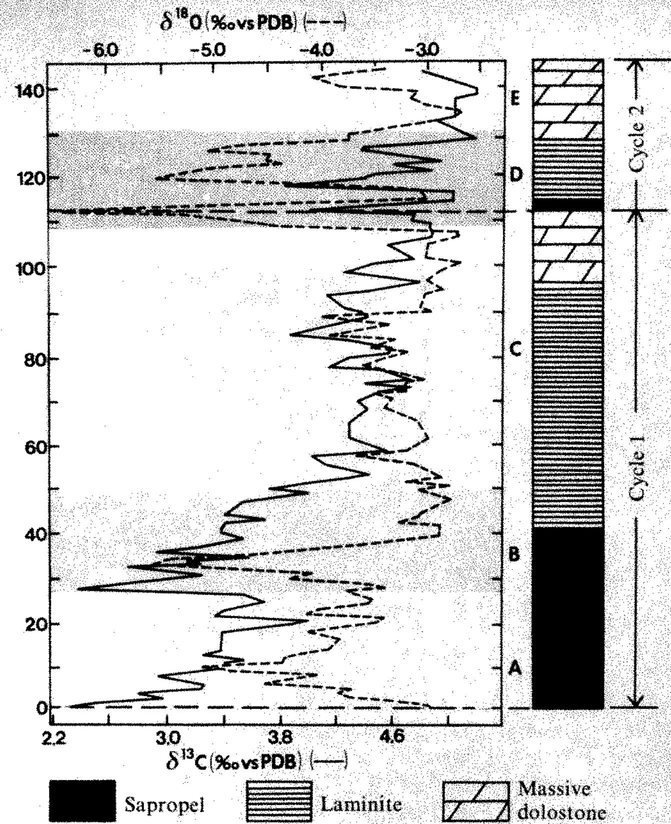


Fig. 1 Carbon and oxygen isotope composition of VT8 core dolomites. Note the spikes of low values of δ^{18}O which occur twice in zone A and zone B representing freshwater reflux of the Zechstein basin. δ^{13}C increases rapidly in zone B and remain constantly high in the upper part of the Marl Slate (zones C–E).

Our understanding of the processes which cause the shift in isotopic ratios will be greatly improved by detailed studies of the transition zone in which the change in ^{13}C is recorded. Magaritz *et al.*[13] noted the presence of the transition zone at the base of the Zechstein sequence in the Marl Slate and its German equivalent the Kupferschiefer. We now discuss the results of a detailed investigation of this isotopic transition zone within the Marl Slate of North-East England.

The Marl Slate represents the earliest extensive marine transgression into the Rotliegend inland drainage basin whose floor lay well below the contemporaneous sea level in Permian times[21]. The preservation of unconsolidated dunes beneath the Marl Slate indicates that the Zechstein transgression was rapid, forming an inland sea up to 250 m deep in places[22]. The Marl Slate is usually less than 1-m thick and mainly comprises organic-rich, dolomitic laminites and closely resembles much younger sapropelic sediments in which lack of bioturbation due to the absence of benthic fauna is a characteristic feature. The present results come from a single core (VT8) drilled by the National Coal Board off the coast of North-East England east of Seaham Harbour (NZ 486 509). This core shows two cycles

0028–0836/82/220389—02$01.00

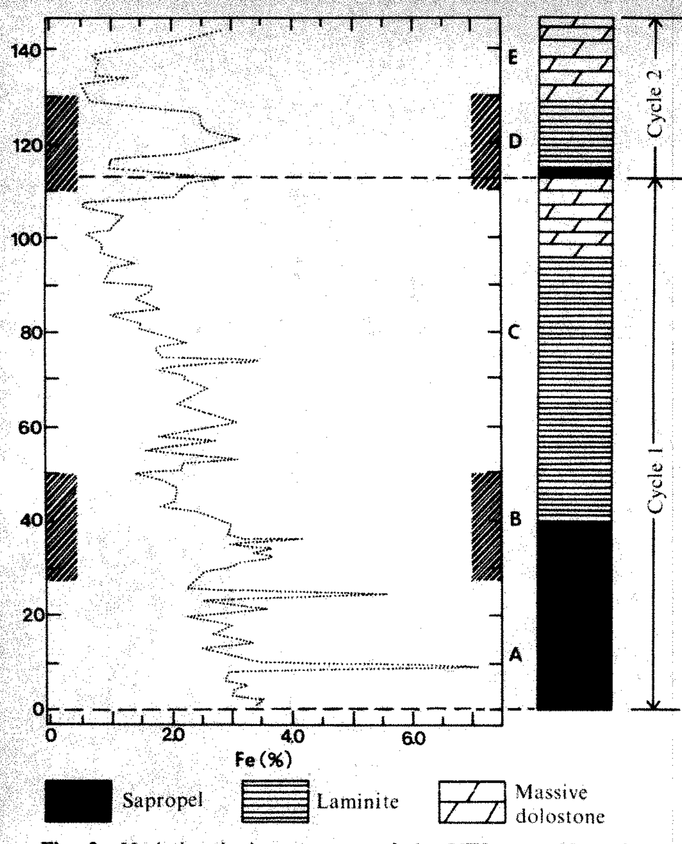


Fig. 2 Variation in iron content of the VT8 core. Note the similarity between the $\delta^{18}O$ curve and iron content.

of sedimentation which include the sequence: sapropel → dolomitic laminite → finely crystalline dolostone (Fig. 1). The core was divided into 144 1-cm thick segments which were used for isotopic and chemical measurements. The carbonate isotopic data (Fig. 1) indicate (1) the lack of correlation between $\delta^{18}O$ and $\delta^{13}C$ values; (2) a gradual enrichment in ^{13}C from the bottom (+2.3‰) to the top (+5.1‰); (3) a $\delta^{18}O$ pattern which exhibits spikes of very low values; (4) a very clear correspondence between the lithological variation and changes in $\delta^{18}O$, a feature which indicates that the isotopic changes are original and not the result of later recrystallization. The preservation of such detailed structure in the isotopic record of the Marl Slate is consistent with its deposition in an anoxic non-bioturbated environment. The organic-rich laminae are interpreted as seasonal phytoplankton blooms and thus enable assessment of the time scale of the events recorded in the core. Oelsner[23] counted the laminae in the Kupferschiefer and arrived at an age of 17,000 yr. Our observations are consistent with this suggesting an overall sedimentation of $\sim 0.08\ \mathrm{mm\ yr^{-1}}$. This value is close to the rate estimated (0.05–$0.07\ \mathrm{mm\ yr^{-1}}$) for the equivalent unit (the Stinkschiefer) in the second Zechstein cycle[24] but greater than the rate of sedimentation of anoxic sediments in the Lower Cretaceous North and South Atlantic ocean which is estimated[25] at $0.01\ \mathrm{mm\ yr^{-1}}$. Our overall estimate of the time represented by the VT8 core is 18,000 yr and the large change in $\delta^{13}C$ values took place within this time interval. The core section can be divided into five macrosegments on the basis of the isotopic variation (Fig. 1), and the main carbon isotope shift occurs in the 22 cm of segment B. This is probably equivalent in time to ~3,000 yr. Even shorter period events are detected in the core in both the oxygen and carbon isotopic record. The shift of more than 1‰ in $\delta^{13}C$ values which occurs within 3 cm (~400 yr) at the bottom of zone A is larger than any shift reported in the Cretaceous[9]. The $\delta^{18}O$ spikes, which are seen in several parts of the core occur within depositional periods of < 5 cm, represent flooding of the basin by freshwater. The period of flooding is similar to that recorded in the Pleistocene sapropelic sediments of the eastern Mediterranean[26]. The low $\delta^{18}O$ values coincide with enrichment of iron (as pyrite) in the sediment (Fig. 2). This correlation indicates that the formation of anoxic conditions was associated with freshwater flooding of the basin. We take this to indicate that in these conditions there was an abundant supply of iron which enhanced iron suphide formation.

The gradual change in $\delta^{13}C$ values in zone B and the early cycle of changes in zone A are considered to represent a change in the dissolved carbonate in the Zechstein Sea. From that point on the $\delta^{13}C$ values of the carbonates in the higher Zechstein cycles remain highly enriched in ^{13}C (refs 13, 14, 19). Such enrichment in ^{13}C is also seen in basins of Permian age other than the Zechstein in various parts of the world[16–18].

We argue that this change in the isotopic composition of the dissolved carbonate in the Zechstein Sea must reflect a similar change on an ocean wide scale. This is because a continuous supply of dissolved salts is needed to account for the succeeding thick carbonate/evaporite sequence. There is no obvious alternative supply of dissolved materials in view of the fact that the other source of dissolved carbonate (freshwater reflux) would result in low $\delta^{13}C$ values. This is evidenced by the less positive $\delta^{13}C$ values seen in the zones of $\delta^{18}O$ minima (Fig. 1).

It is not clear whether all the water column of the Permian ocean (all the oceanic bicarbonate reservoir) changed its carbon isotopic composition (reflecting a large variation in the C–S–N cycle) or whether the change occurred only in the top oceanic layer, because of the lack of a deep oceanic sequence for comparison. The size of the shift in $\delta^{13}C$ values in the Permian period is at least twice the size of any shift in the Cretaceous. Simple mass isotope calculations show that a shift of 3‰ in $\delta^{13}C$ will correspond to almost double the amount of organic material removed from the system at a specific time. Enrichment of 0.7‰ in $\delta^{13}C$ was estimated for the mean dissolved carbonate in the ocean water in a period < 10,000 yr for the Pleistocene glacial/interglacial transition[27]. The shift recorded in the Permian probably took place at a similar rate, but the effect was tripled (~2.3‰).

The $\delta^{34}S$ values of the Zechstein sulphates are all depleted in ^{34}S (ref. 28) which indicates that the accompanying effect in the sulphur cycle had already taken place by the end of the first Zechstein cycle. The very widespread regression ending the Permian Period and the Zechstein Sea was followed by the return to normal $\delta^{13}C$ values in early Triassic carbonates[29].

Received 20 November 1981; accepted 31 March 1982.

1. Garrels, R. M. & Perry, F. A. in *The Sea* Vol. 5 (ed. Goldberg, E. D.) 303–336 (Wiley, New York, 1975).
2. Mackenzie, F. T. & Pigott, J. D. *J. geol. Soc. Lond.* **138,** 183–196 (1981).
3. Schidlowski, M., Junge, C. E. & Pietrek, H. *J. geophys. Res.* **82,** 2557–2565 (1977).
4. Jeffery, P. M., Compston, W., Greenhalch, D. & Debacter, Y. *Geochim. cosmochim Acta* **7,** 255–286 (1955).
5. Weber, J. *Geochem. cosmochim Acta* **31,** 2343–2351 (1967).
6. Shackleton, N. J. in *The Fate of Fossil Fuel CO_2 in the Ocean* (eds Anderson, N. R. & Malahoff, A.) 543–563 (Plenum, New York, 1977).
7. Fischer, A. G. & Arthur, M. A. *SEPM. Spec. Publ.* **25,** 19–50 (1977).
8. Boersma, A. & Shackleton, N. J. *Cretaceous–Tertiary Boundary Event Symp.* **2,** 40–54 (1979).
9. Scholle, P. A. & Arthur, M. A. *Bull. geol. Soc. Am.* **64,** 67–87 (1980).
10. Kroopnick, P. M., Margolis, S. V. & Wang, C. S. in *The Fate of Fossil Fuel CO_2 in the Ocean* (eds Anderson, N. R. & Malahoff, A.) 295–321 (Plenum, New York, 1977).
11. Jenkyns, H. C. *J. geol. Soc. Lond.* **137,** 171–181 (1980).
12. Veizer, J., Holser, W. T. &Wilgus, C. K. *Geochim. cosmochim Acta* **44,** 579–588 (1980).
13. Magaritz, M., Turner, P. & Kading, K-Ch. *Geol. J.* **16,** 243–254 (1981).
14. Magaritz, M. & Schulze, K-H. *Cont. Sedim.* **9,** 269–277 (1980).
15. Ponov, A. B. *Geokimiya* **8,** 1252–1277 (1976).
16. Compston, W. *Geochim. cosmochim. Acta* **18,** 1–22 (1960).
17. Dean, W. E., Davies, G. R. & Anderson, R. Y. *Geology* **3,** 367–372 (1975).
18. Osaki, S. *Geochim. J.,* **6,** 463–477 (1973).
19. Botz, R. thesis, Univ. Heidelberg (1979).
20. Thode, H. G., Monster, J. & Dunford, H. B. *Bull. Am. Ass. petrol. Geol.* **42,** 2619–2641 (1958).
21. Smith, D. B *J. geol. Soc. Lond.* **136,** 155–156 (1979).
22. Smith, D. B. *Contr. Sedim.* **9,** (1980).
23. Oelsner, O. *Freiberg. Forsch. C.* **58,** 106–113 (1959).
24. Brongersma-Sanders, M. *Mar. Geol.* **11,** 123–144 (1971).
25. Arthur, M. A. & Natland, J. H. in *Results of Deep Drilling in the Atlantic Ocean* Vol. 3, 375–401 (U.S. Govt. Printing Office, Washington DC, 1979).
26. Ryan, W. B. F. & Cita, M. B. *Mar. Geol.* **23,** 197–215 (1977).
27. Broecker, W. S. in *International Conference on Evolution of Planetary Atmospheres and Climatology of the Earth,* 165–190 (Cent. Nat. Etud. Spat. (France), 1978).
28. Holser, W. T. & Kaplan, I. R. *Chem. Geol.* **1,** 93–135 (1966).
29. Wilgus, C. K. thesis, Univ. Oregon, (1981).

Hotspots, polar wander, Mesozoic convection and the geoid

Don L. Anderson

Seismological Laboratory, California Institute of Technology, Pasadena, California 91125, USA

The geoid bears little relation to present tectonic features of the Earth other than trenches and hotspots. The Mesozoic supercontinent of Pangea, however, apparently occupied a central position in the Atlantic–African geoid high. This and the equatorial Pacific geoid high contain most of the world's hotspots[1,2]. The plateaus and rises which are now in the western Pacific formed in the Pacific geoid high and this may have been the early Mesozoic–late Palaeozoic position of a large part of Asia and other fragments of the Pacific rim continents. The major global geoid highs were regions of extensive Cretaceous volcanism and may be the former sites of continental aggregations and mantle insulation and, therefore, hotter-than-normal mantle. The pent-up heat causes rifts and hotspots and results in uplift, magmatism, fragmentation and dispersal of the continents and the subsequent formation of plateaus, aseismic ridges and seamount chains which cause a global rise in sea-level. Convection in the upper mantle caused by such lateral temperature gradients is intrinsically episodic. A geoid anomaly of 50 m can be formed in about 100 Myr by continental insulation. We show here that such geoid anomalies are long-lived and may be used to remove the ambiguity in early Mesozoic–late Palaeozoic plate reconstructions. Geoid highs control the rotation axis of the Earth and, in effect, bring long-lived continental aggregations to the Equator. Many aspects of continental geology such as vertical-tectonics and episodicity of magmatism and transgressions can be explained by continental insulation.

The Earth's largest positive geoid height anomalies are associated with subduction zones and hotspots and bear no simple relationship to other present-day tectonic regions such as continents and ridges. When the subduction-related geoid highs are removed from the observed field the residual geoid shows broad highs over the central Pacific and the eastern Atlantic–African regions[1,2]. Like the total geoid, the residual geoid does not reflect the present-day distribution of continents and oceans and shows little trace of the ocean ridge system. Geoid highs, however, correlate with hotspots[1,2] and with regions of anomalously shallow ocean floor and sites of extensive Cretaceous volcanism.

The Atlantic–African geoid high extends from Iceland through the North Atlantic and Africa to the Kerguelen plateau and from the middle of the Atlantic to the Arabian penninsula and western Europe. Most of the Atlantic, Indian Ocean, African and European hotspots are inside this anomaly. Iceland, Trindade, Tristan, Kerguelen, Reunion, Afar, Eiffel and Jan Mayan form the 20-m boundary of the anomaly and seem to control its shape. The Azores, Canaries, New England seamounts, St Helena, Crozet and the African hotspots are interior to the anomaly. This, plus similar evidence in the Pacific, suggests that geoid highs may be associated with hotter-than-normal mantle.

Although the geoid high cuts across present-day ridges and continents there is a remarkable correspondence of the predrift assemblage of continents with the present location of both the geoid anomaly and hotspots (Fig. 1). Reconstruction of the mid-Mesozoic configuration of the continents[3] reveals, in addition, that most of the large shield areas of the world are contained inside the Atlantic–African geoid high. Most of the Phanerozoic platforms are also in this area, the main exceptions being the Siberian and South-East Asian platforms.

The area inside the geoid high is also characterized by higher-than-normal present elevations, for example, in Africa, the North Atlantic[4] and the Indian Ocean south-east of Africa[4]. This also holds true for the axial depth of oceanic ridges[5]. Most of the continental areas were above sea level from the Permo-Carboniferous to the Triassic, at which time there was subsidence in eastern North and South America, central and southern Africa, Europe and Arabia[6]. The widespread uplift, magmatism, breakup and initial dispersal of the Pangean landmass apparently occurred while the continents were centrally located with respect to the present geoid anomaly. The subsequent motions of the Atlantic bordering plates were largely directed away from the anomaly. This suggests that the residual geoid high, hotspots, the distribution of continents during the late Palaeozoic and early Mesozoic and their uplift and subsequent dispersal and subsidence are all related. The shields are regions of abnormally thick lithosphere. The thickest lithosphere is in eastern and central North America, northeastern South America, northwestern and central Africa and northern Siberia[7]. These regions were all within the area of Fig. 1 at 200 Myr.

The area in Fig. 1 also experienced exceptional magmatism during the Mesozoic. The great flood basalts of Siberia and South Africa were formed during the Triassic and Jurassic, possibly at the sites of the Jan Mayen and Crozet hotspots[3]. The plateau basalt provinces of South-east Greenland and Brazil were formed during the Cretaceous and early Cenozoic, possibly at the sites of the Iceland and Tristan hotspots[3]. The Walvis Ridge, the Rio Grande Rise and the plateaus in the western Indian Ocean are mainly on Mesozoic and Palaeocene crust. A large part of the Pacific also experienced extensive on- and off-ridge volcanism in the Cretaceous[8,9]. This extensive ridge and hotspot volcanism may have caused the rapid rise in sea level during the Jurassic and Cretaceous[9]. If sea level[10] can be used as a guide to the volume of the ocean basins then from the end of the Cretaceous to the end of the Oligocene was a period of less intense oceanic volcanism and subsidence of the oceanic and continental crust. Sea-level variations may therefore indicate that the thermal and geoid anomalies formed in the Palaeozoic and have attentuated since the early Mesozoic.

Much of the Pacific rim appears to be terrain which originated in the Pacific[11,12]. The possibility of a continent centrally located in the Pacific has been discussed for some time[13], but its location has been an enigma and its size uncertain. The central Pacific geoid anomaly, like the Atlantic–African anomaly, may mark the early Mesozoic or late Palaeozoic location of a large continental landmass and, later, the site of extensive Cretaceous ridge crest and midplate volcanism.

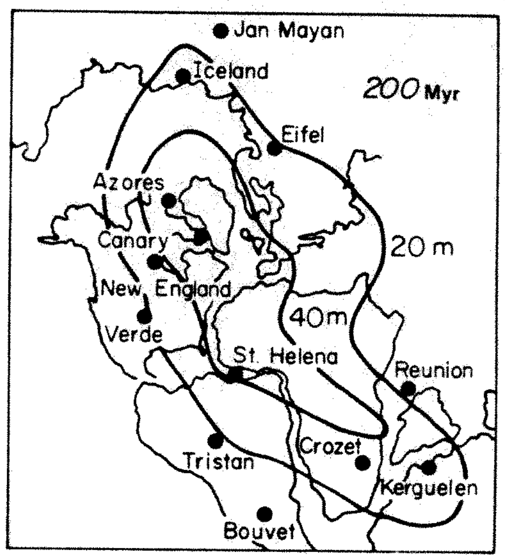


Fig. 1 The Jurassic configuration of the continents, based on Morgan's hotspot-based reconstruction[3], superposed on the Atlantic–African residual geoid high[1,2]. Note that most of the stable shield areas are inside the 20-m contour. Selected hotspots (●) are also shown.

0028-0836/82/220391—03$01.00

Palaeomagnetic data indicate that various blocks of Asia such as Kolyma, Sikhote Alin, Sino-Korea, Yangtze, South-east Asia and Japan have moved northwards by up to 32° since the Permian[11]. It is unlikely that they were in the vicinity of Australia or associated with Gondwana[11] and a central Pacific location is likely. Part or all of Alaska and northwestern North America were also far south of their present positions, relative to North America, in the Permian[12]. The same may be true of California, Mexico, Central America and other accreted terrain in the Pacific rim continents.

The central Pacific residual geoid high (>20 m) extends from Australia to the East Pacific Rise and from Hawaii to New Zealand. It encompasses most of the Pacific hotspots and is approximately antipodal to the Atlantic–African anomaly.

The western Pacific contains numerous plateaus, ridges, rises and seamounts which have been carried far to the north-west from their point of origin. The Ontong–Java plateau for example, presently on the Equator, was formed at 33–40°S in the mid-Cretaceous[14]. The Hess rise, Line Islands ridge and Necker ridge were formed in the Cretaceous on the Pacific, Farallon and Phoenix ridge-crests[8] which were, at the time, in the eastern Pacific in the vicinity of the polynesian seamount province. This ridge-crest volcanism was accompanied by extensive deep-water volcanism[8,9] and, possibly, rapid seafloor spreading[15]. The Caribbean and Bering Seas may have formed at the same location and carried northward on the Farallon and Kula plates. It is significant that the Caribbean and the anomalous regions in the Pacific have similar geophysical and geochemical characteristics[16].

The Pacific geoid high also has anomalously shallow bathymetry[4,5]. This shallow bathymetry, the 'Darwin Rise', extends from the East Pacific Rise to the north-west Pacific in the direction of plate motion, and includes the central and south-west Pacific hotspots. The extensive volcanism in the central western Pacific between ~70 and 120 Myr (refs 8, 9, 16) occurred about 60°–100° to the south-east of its present position, in the hotspot frame. This would place the event in the vicinity of the southeastern Pacific hotspots and the eastern part of the geoid anomaly. This suggests that the anomaly dates back to at least early Cretaceous. A similar thermal event in the Caribbean[17] may mean that it was also in this region in the Cretaceous, particularly since the basalts in the Caribbean are similar to those in the western Pacific[16].

Going back even further are the Triassic basalts in Wrangellia[11] of northwestern North America and the Permian greenstones in Japan[13], both of which formed in equational latitudes and subsequently drifted to the north and north-west respectively. We suggest that they also formed in the Pacific geoid high.

Plate reconstructions based on palaeomagnetism or palaeoclimatology have arbitrary longitudinal relationships between the plates. Figure 2 gives a possible late Palaeozoic reconstruction which places Africa, South America, Antarctica and Europe in the Atlantic–African geoid high and Asia in the Central Pacific geoid high. This is similar to previous reconstructions[6,18] in that North America is adjacent to western South America but Asia is placed much closer to North America. The biogeographic similarities between Asia and western North America are well known[19] and the placement of North America provides the continuity between the major land masses which is suggested by some palaeontological data. With this reconstruction Gondwanaland and Asia were the continental aggregations that caused the geoid anomalies and they brought themselves to the Equator, by the Goldreich–Toomre mechanism[20], before their breakup and dispersal. Parts of Alaska, western North America and West Antarctica may also have migrated away from the Pacific geoid high.

We propose that episodicity in continental drift, polar wander, sea-level variations and magmatism are due to the effect of thick continental lithosphere on convection in the mantle and that supercontinents 'insulated' large parts of the mantle for more than 10^8 yr. The excess heat caused uplift by thermal expansion and partial melting and, eventually, breakup and dispersal of the supercontinents. A large geoid high is generated by this expansion and, if this is the dominant feature of the geoid, the spin axis of the Earth would change so as to centre the high on the Equator[20], much as it is today. The continents stand high while they are within the thermal and geoid high and subside as they drift over cooler mantle.

The association of the Atlantic–Africa geoid high with the former position of the continents and the plateau basalt provinces with currently active hotspots suggests that Mesozoic convection patterns in the mantle still exist, as proposed by Menard.[21]. The mid-Atlantic and Atlantic–Indian ridges and the Atlantic, African and Indian ocean hotspots and the East African rift are regions where heat and material is being efficiently removed from the uppermantle in the Atlantic–African geoid high. The East Pacific rise and the Pacific hotspots are serving the same role in the Pacific geoid high. If geoid anomalies and hotspots are due to a long period of continental insulation and if these regions had higher-than-normal heat flow and magmatism for the past 100 Myr, then we might expect that these features will wane with time.

Global isostatic geoid anomalies can be written[22,23]

$$N \cong \frac{2\pi G}{g}\frac{(l+2)}{(2l+1)}\rho h \delta h$$

where G is the gravitational constant, g is the acceleration of gravity, l is degree of spherical harmonic, ρ is density, h is the depth of the anomalous region and δh is the elevation anomaly associated with the mass anomaly. For a coefficient of thermal expansion of $3\times10^{-5}\,°C^{-1}$, a temperature excess of 70 °C in the upper 500 km of the mantle gives $\delta h = 1$ km and $N = 50$ m for $l = 3$. A similar effect can be achieved by 1% melting. Some combination of heating and melting is likely.

The average heat loss through oceanic areas is $2.4\ \mu cal\ cm^{-2}\ s^{-1}$ (ref. 24). Continental heat flux is about 0.6 of this in spite of the higher radioactivity of continental crust. If we assume that the oceanic heat flux is available everywhere and that there is a balance between heat production and heat loss under oceans the temperature rise under an extended stationary continent is 0.60 °C Myr^{-1} for a 500-km thick column or, alternatively, 6% melting, using a latent heat of 100 cal g^{-1}. Thus, 100 Myr of continental insulation can give the observed geoid and elevation anomalies. The continental aggregations were relatively stable for at least this long[6,18]. True polar wander is likely while the geoid anomalies are developing.

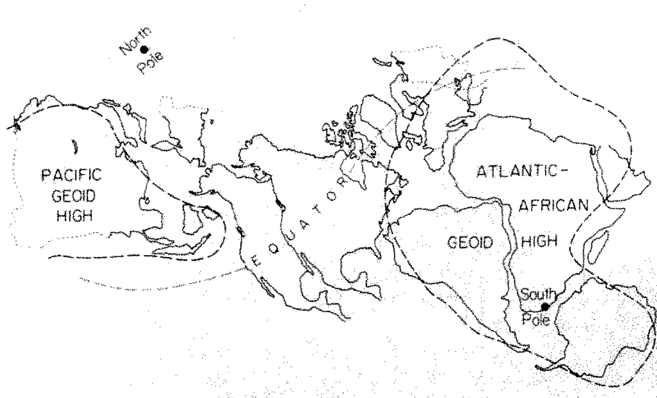


Fig. 2 Possible late Palaeozoic continental configurations relative to the current residual geoid highs and the Palaeoequator. This is similar to previous reconstructions[6,18] except that the Pacific geoid high is used to position eastern Asia. Australia and India are not shown. The dashed lines enclose regions of high elevation (>20 m) of the residual geoid[1,2]. If the geoid highs were generated under Gondwana and Asia they formed in high latitudes and were subsequently brought to the Equator by their affect on the rotation axis of the Earth. The geoid highs are currently centred on the Equator. Parts of western North America may also have been in the Pacific geoid high. This and the eastern migration of North America, relative to South America, during the Palaeozoic are indicated schematically by plotting North America in two positions. Asia is probably also fragmented at this time[11].

After the continents move off the mantle which they have been insulating the thermal and geoid anomalies decay by convection in the mantle and conduction through the newly-formed oceanic lithosphere. The rate of temperature decrease, $\dot{\theta}$, can be written

$$\dot{\theta} = \frac{-\kappa\theta}{h^2}\left(\frac{R}{R_{cr}}\right)^{1/3}$$

where κ is the thermal diffusivity, R is the Rayleigh number and R_{cr} is the critical Rayleigh number. For $R_{cr} = 10^3$, $R = 10^6$, $\kappa = 10^{-2}\ \mathrm{cm^2\ s^{-1}}$, $\theta = 10^3$K and $h = 500$ km

$$\dot{\theta} = -0.12\ \mathrm{°C\ Myr^{-1}}$$

This gives a decrease in elevation of 180 m in 100 Myr. The corresponding geoid height reduction is ~9 m. Thus, geoid anomalies caused by continental insulation are long-lived features.

Horizontal temperature gradients can drive continental drift[25-28]. The velocities decrease as the distance increases away from the heat source and as the thermal anomaly decays. The geoid high also decays. Thick continental lithosphere then insulates a new part of the mantle and the cycle repeats. Periods of rapid polar motion and continental drift follow periods of continental stability and mantle insulation.

The relatively slow motion of the continents[29] or the pole[30], or both, during the Permian and the relative stability of sea level during this period suggest that the geoid anomalies were developing at this time. Continental drift velocities[29] and sea level[10] changed rapidly during the Triassic and Jurassic. This we interpret as the start of extensive rift and hotspot magmatism and the decay of the thermal and geoid anomalies.

The rotation axis was apparently stable from ~200 to 80 Myr (ref. 3) but has shifted with respect to the hotspot reference frame for at least the past 65 Myr (ref. 31). In the framework of our model this shift would represent a growth of the equatorial region of the Atlantic–African high due to continued insulation by Africa, a decay of the South Pacific or North Atlantic portions of the highs due to extensive Cretaceous volcanism, a reconfiguration of the world's subduction zones or some combination.

It is difficult to separate true polar wander and the apparent wander due to continental drift. In the present model they are intimately related. If ridge volcanism is precluded for an extensive period of time, as seems to have been the case in the mantle under Pangea during the early Mesozoic and late Palaeozoic, the mantle will warm up and/or partially melt. In either case a geoid anomaly will develop. This anomaly will affect convection and plate motions in a manner previously suggested[25,32] and will also affect the rotation axis of the Earth[20]. As the thermal anomalies seem to be long-lived and it takes a long time to establish a new thermal regime, major shifts of the rotation axis due to this mechanism will be infrequent and gradual.

Goldreich and Toomre[20] have shown that density inhomogeneities caused by convection in the mantle will control the orientation of the spin axis and that large shifts of the pole, that is, true polar wander, are to be expected in a dynamic Earth. The present orientation of the pole apparently bears no relationship to the present distribution of continents, ridges or trenches[32]. On the other hand, the geoid bears a close relationship to the present distribution of hotspots and to the former distribution of continents and areas of extensive Cretaceous ridge and plateau volcanism.

If our conjecture about the relationship between geoid highs and the former position of continents is correct, then we would conclude that there were two large antipodal continental aggregations in the Permian. The antipodal supercontinents were presumably formed by a previous episode of continental drift which swept buoyant material away from an earlier configuration of thermal and geoid highs. Using sea level as a guide, with high sea-level indicating rifting and extensive sea floor volcanism, the supercontinents were stable from 300 to 150 Myr ago and the previous continental assemblages were dispersing between 500 and 350 Myr. The suggestions that hotspots may be a primary cause of continental breakup[3] and that they affect polar wander[1,2] are not new. What is new is the suggestion that continents cause geoid anomalies and hotspots and, indirectly, control the rotation axis of the Earth.

I thank Brad Hager, Clem Chase, Kevin Burke and Henri Nataf for reviewing the manuscript, and Lynn Sykes for stimulating discussions. Art Boucot gave useful advice on biogeography. Brad Hager showed me how to calculate global isostatic geoid anomalies. This work was supported by NASA grant NSG-7610. Contribution no. 3710, Division of Geological and Planetary Sciences, California Institute of Technology.

Received 16 November 1981; accepted 27 January 1982.

1. Chase, C. *Nature* **282**, 464–468 (1979).
2. Crough, S. & Jurdy, D. *Earth planet. Sci. Lett.* **48**, 15–22 (1980).
3. Morgan, W. J. in *The Sea* Vol. 7 (ed. Emiliani, C.) 443–488 (Wiley-Interscience, New York 1981).
4. Cochran, U. & Talwani, M. *Geophys. J.* **50**, 495–552 (1977).
5. Menard, H. & Dorman, L. *J. geophys. Res.* **82**, 5329–5335 (1977).
6. Kanasewich, E. R., Havskov, J. & Evans, M. *Can. J. Earth Sci.* **15**, 919–955 (1978).
7. Pollack, H. & Chapman, D. S. *Tectonophysics* **38**, 279–296 (1977).
8. Watts, A. B., Bodine, V. & Ribe, N. *Nature* **283**, 532–537 (1980).
9. Schlanger, S. O., Jenkyns, H. C. & Premoli-Silva, I. *Earth planet. Sci. Lett.* **52**, 435–449 (1981).
10. Vail, P., Mitchum, R. M. Jr & Thompson, S. III *Am. Ass. petrol. Geol. Mem.* **26**, 83–97 (1978).
11. McElhinny, M., Embleton, B. J. J., Ma, X. H. & Zhang, Z. K. *Nature* **293**, 212–216 (1981).
12. Panuska, B. & Stone, D. *Nature* **293**, 561–563 (1981).
13. Hattori, I. & Hirooka, K. *Tectonophysics* **57**, 211–235 (1979).
14. Hammond, S., Kroenke, L., Thayer, F. & Keelong, D. *Nature* **255**, 46–47 (1975).
15. Larson, R. & Chase, C. *Bull. geol. Soc. Am.* **83**, 3627–3644 (1972).
16. Batiza, R., Larson, R., Schlanger, S., Scheka, S. & Tokuyama, H. *Nature* **286**, 476–478 (1980).
17. Burke, K., Fox, D. & Senqor A. M. *J. geophys. Res.* **83**, 3949–3954 (1978).
18. Smith, A. G., Hurley, A. M. & Briden, J. C. *Phanerozoic Paleocontinental World Maps* (Cambridge University Press, 1981).
19. Boucot, A. J. & Gray, J. in *Historical Biogeography, Plate Tectonics, and the Changing Environment* (eds Gray, J. & Boucot, A. J.) 465–482 (Oregon State University Press, 1979).
20. Goldreich, P. & Toomre, A. *J. geophys. Res.* **74**, 2555–2567 (1969).
21. Menard, H. W. *Earth Planet. Sci. Lett* **20**, 237–241 (1973).
22. Dahlen, F. A. (in preparation).
23. Hager, B. H. (in preparation).
24. Sclater, J. G., Parsons, B. & Jaupart, C. *J. geophys. Res.* **86**, 11,535–11,552 (1981).
25. Elder, J. *Nature* **214**, 657–660 (1967).
26. Ichiye, T. *J. geophys. Res.* **76**, 1139–1153 (1971).
27. Busse, F. *J. Geophys. J. R. astr. Soc.* **52**, 1–12 (1978).
28. Froidevaux, C. & Nataf, H. *Sonderdruck aus der Geologischen Rundschau Band* **70**, 166–176 (1981).
29. Gordon, R. G., McWilliams, M. O. & Cox, A. *J. geophys. Res.* **84**, 5480–5486 (1979).
30. Van Alstein, D. R. thesis, California Inst. Technol. (1979).
31. Gordon, R. & Cape, C. *Earth planet. Sci. Lett.* **55**, 37–47 (1981).
32. Jurdy, D. *J. geophys. Res.* **83**, 4989–4994 (1978).

Development of earthquake-induced fissures in the Main Ethiopian Rift

Laike M. Asfaw

Geophysical Observatory, Addis Ababa University, PO Box 1176, Addis Ababa, Ethiopia

A series of episodes involving land subsidence along an elongated formation of pits connected by fissures have been observed in various parts of the Main Ethiopian Rift[1-4] over the past 25 yr. Although other workers[1] have proposed a possible relationship between tectonics, subsurface drainage and subsidence no field observations have related earthquakes to such development of fissures. I present here the first observation of the development of fissures following an earthquake swarm with local magnitude, $M_L \leqslant 4$ in the Main Ethiopian Rift north of the Fentale Mountain (Fig. 1) between January and March 1981 (ref. 5). This showed that there could be a variety of developments in the evolution of earthquake-induced fissures depending on geological formation, vegetation cover and drainage; the implication on similar earlier observations is considered.

That the Afar Depression and the Main Ethiopian Rift are regions under tension is well established through geodetic

0028-0836/82/220393—03$01.00

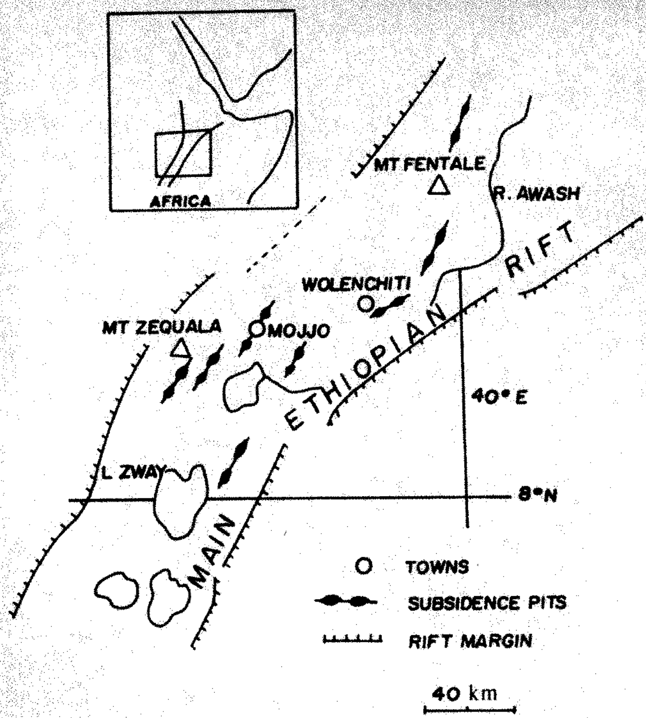


Fig. 1 Sites of subsidence pits associated with fissures in the Main Ethiopian Rift including the event of January–March 1981 north of the Fentale Mountain.

measurements and other surface observations[6,7]. The surface observations consist of tensional fissures and faults along the three major rift trends of the Red Sea, Gulf of Aden and the Main Ethiopian Rift[8,9]. In the Main Ethiopian Rift the Wonji Fault Belt is characterized by fissures and faults[10] and sometimes fissures can be identified by elongated green belts of vegetation in this rift. Here the sediment deposit, rainfall and vegetation make the development of fissures more complicated and varied compared with those in the more arid rifts in the Afar Depression trending in the Red Sea and the Gulf of Aden directions.

Earlier investigators[1-4] had noted subsidence and erosion associated with fissures in the Main Ethiopian Rift. In most of the cases the features observed consist of subsidence pits connected by fissures and oriented either obliquely or parallel to the rift trend, generally indicating a relation to the tectonics of the region. Gouin and Mohr[1] tried to relate earthquakes which occurred a few months earlier in 1966 with such subsidence in the rift floor; they suggested a possible relationship between tectonics, underground drainage and surface subsidence, however, neither a direct relation to earthquakes nor the actual development of the fissures was observed. All reports of subsidences associated with these type of formations were made after the incidences, usually following heavy rains, and therefore the problem of their origin was not solved.

The occurrence of fissures during an earthquake swarm in the Main Ethiopian Rift north of the volcanic centre of Fentale between January and March 1981 and the subsequent erosion provided the first opportunity to follow the development and identify the main features in such a process. In the epicentral area fissures, 2 cm wide, run on sediments and basalt flows for a few kilometres and had a NNE–SSW orientation following the rift trend. After an interval of 6 months, which included a rainy season, the fissures that occurred on sediments on the site were found to have changed considerably while fissures on basalt flows showed very little changes.

Fissures in the sedimentary layer became surface and subsurface drainage channels and the features that developed during a subsequent erosion were determined by the characteristics of sites such as the vegetation cover and the exact nature of geological formations. Depending on the vegetation cover the erosion of fissures on sedimentary formation was either subsurface, where there was deep-rooted vegetation, or gully type, where there was no substantial growth (see Figs 2*b* and 3*c*). In the case of subsurface erosion of fissures the subsidence of the thin sedimentary veneer, which was held together by roots of vegetation, following heavy rains, resulted in the sporadic occurrence of pits connected by fissures (see Figs 2*c* and 3*b*). On the other hand, fissures on sediments with flat topography and therefore with minimum surface drainage were sealed off and did not develop further (see Figs 2*a* and 3*a*). The observed states of erosion following the earthquake swarm between January and March can be classified into three main categories as indicated in Fig. 2.

The features in the third type of development, Fig. 2*c*, bear a strong resemblance to many other observations of subsidences in the Main Ethiopian Rift. The depths of such formations were much larger than their widths and at times full-grown trees had been swallowed in the pits. The great depths attained compared with their widths can be explained by erosion due to subsurface drainage which would mainly be conditioned by the water table there. In such situations both the bottom and lower walls of the fissures would be eroded in an unrelated manner to the overlying walls and cover thereby deepening and widening the lower part disproportionately. Usually the pits occur where there is deep-rooted vegetation. Here the roots would fracture the sediment, facilitating the process of erosion underneath—especially laterally—widening the original fissure into pits. Of course, vegetation would minimize surface erosion and not necessarily subsurface erosion. The sporadic occurrence of pits is partly due to the sporadic nature of the distribution of deep-rooted vegetation and partly due to geological factors.

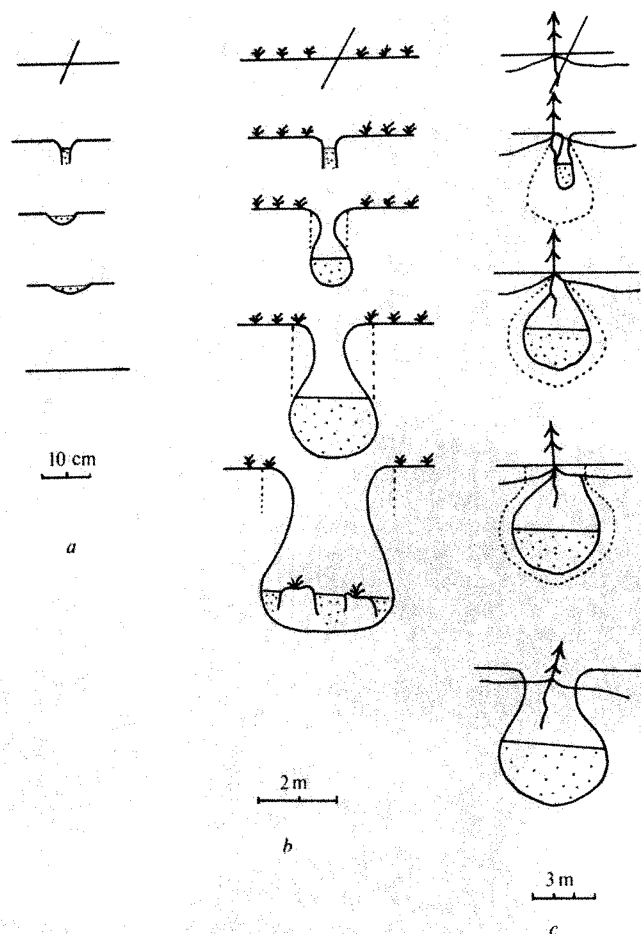


Fig. 2 Types of development of fissures on sediments: *a*, fissures on flat ground with poor drainage; *b* gully type erosion of fissures; *c*, development of subsidence pits. Appropoximate scales do not include vegetation. No vertical exaggeration.

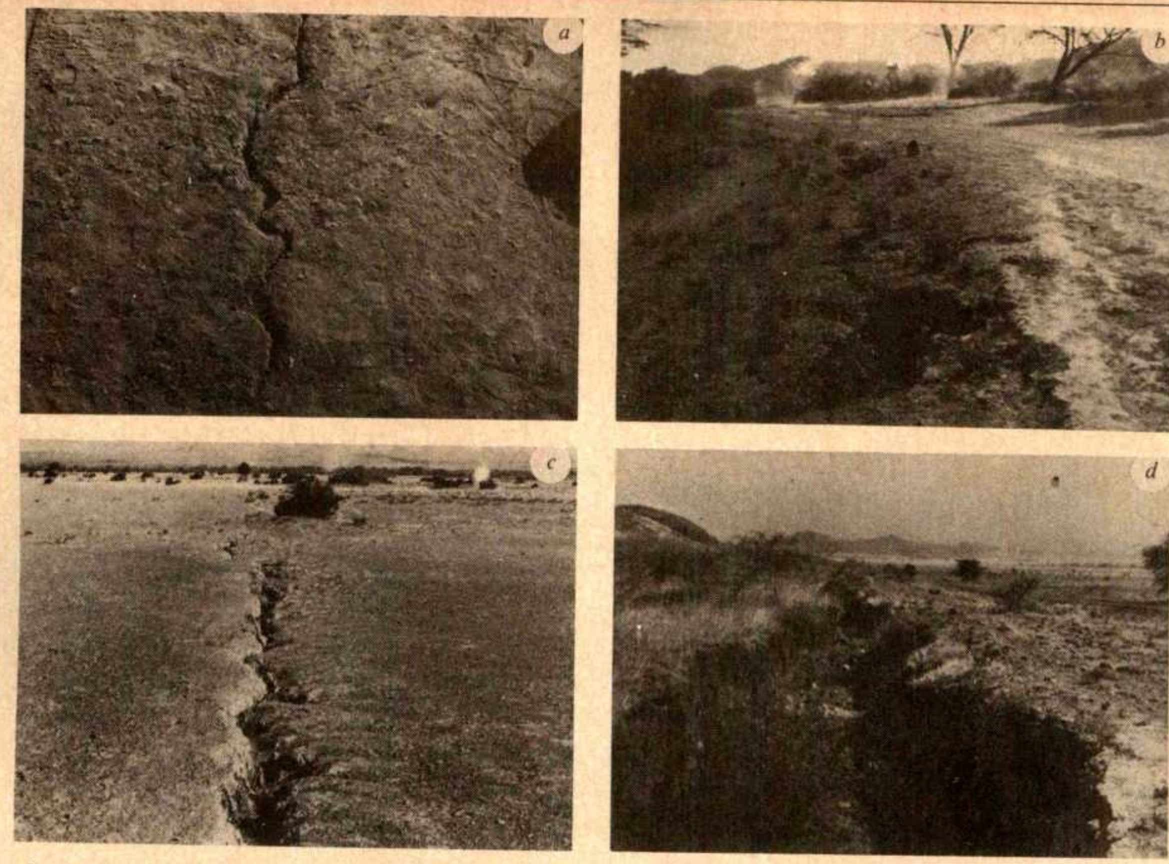


Fig. 3 *a*, Segment of a crack, 2-cm wide, which occurred during the earthquake swarm of January–March 1981 north of Fentale Mountain. This crack was sealed off 6 months later in June 1981. *b*, Subsequent development of pits, about 1 m wide and 50 cm deep, 6 months later, 500 m to the north of Fig. 3*a*. *c*, A small gully type development, 15 cm wide and 10 cm deep, 300 m to the south of Fig. 3*a*. *d*, Gully type erosion of a fissure, 2 m wide and 1.5 m deep, to the west of Fentale Mountain.

The orientation of pits and fissures relative to the rift trend is indicative of their tectonic origin where tension in the region would be associated with fissures oriented generally in the direction of the rift.

The third type of development of fissures, Fig. 3*b*, following the earthquake swarm of 1981, took only 6 months but the main features that characterize subsidence fissures elsewhere had developed. For example, the Mojjo and Wolenchiti fissures and subsidence pits[1–3] could have been considerably older and their rates of development different due to variation in precipitation, subsurface drainage and geological formation. Allowing for age differences the third type of development fully accounts for all the features observed in various parts of the Main Ethiopian Rift. Among other things, the earthquake incidence between January and March 1981 showed that it is possible for earthquake swarms with local magnitude, $M_L \leq 4$ to be associated with fissures. This association is plausible in view of the thin crust there where seismic moments and therefore magnitudes of earthquakes would be small.

On the other hand the segment of the crack shown in Fig. 3*a* was sealed off 6 months later and no trace could be found, while 300 m to the south (Fig. 3*c*) a small gully type development was initiated. These developments correspond to the schemes shown in figs 2*a* and *b* respectively.

Received 20 October 1981; accepted 20 March 1982.

1. Gouin, P. & Mohr, P. *Bull. geophys. Obs.* **10**, 39–46 (1967).
2. *Smithsonian Instn CFSLP Event Card* No. 1019 (1970).
3. *Smithsonian Instn. CFSLP Event Card* No. 1231 (1971).
4. Williams, M. A. J. *Sinet Ethiop. J. Sci.* **3**, 2 (1981).
5. Laike, M. A. *Proc. Conf. Earthquake Swarm* (ECA in the press).
6. Tarantola, A. *et al. Earth planet. Sci. Lett.* **45**, 435–444 (1979).
7. Mohr, P. *et al. Tectonophysics* **44**, 141–160 (1978).
8. Varet, J. & Gasse, F. *Geology of Central and Southern Afar* (CNRS, Paris, 1978).
9. Harrison, C. G. A. *et al.* in *Afar Depression of Ethiopia*, 178–198 (Schweizerbart, Stuttgart, 1975).
10. Mohr, P. *Bull. geophys. Obs.* **11**, 1–65 (1967).

Eruption at Le Piton de la Fournaise volcano on 3 February 1981

P. Bachelery, P. A. Blum, J. L. Cheminee, L. Chevallier, R. Gaulon*, N. Girardin*, C. Jaupart†, F. Lalanne, J. L. Le Mouel, J. C. Ruegg* & P. Vincent

Observatoires Volcanologiques, Institut de Physique du Globe de Paris, Tour 14–24, Université Paris VI, 4 Place Jussieu, 75230 Paris Cedex 05, France

Piton de la Fournaise in the island of La Réunion, Indian Ocean, is a young (0.35-Myr old) basaltic shield volcano built on the flank of the extinct Piton des Neiges volcano. It is one of the most active volcanoes in the world (magma output rate of 0.32 $m^3 s^{-1}$)[1,2] and has been under continuous surveillance by the Volcanological Observatory of the Institut de Physique du Globe de Paris since August 1980. The first detailed observations on the volcano reveal a remarkably low level of background seismicity. The 3 February 1981 eruption was preceded by a deep event of volcanic origin at depths of 10–15 km on 2 January; 20 days of quiescence followed this event, until the onset of shallow seismicity close to the eruptive centres on 22 January. The eruption began after 13 days of sustained shallow seismicity and occurred in three phases over 92 days and produced about 10×10^6 m^3 of lava.

A preliminary report of the activity preceding the eruption is given in ref. 3. The seismic network consists of four vertical seismometers and one three-component seismometer (PHR) located around the caldera (Fig. 1*a*). The instruments have a linear response at frequencies of 1–30 Hz with a gain of 10^3. Signals are transmitted by radio to the observatory where they

* Also at Laboratoire L.A. 195.
† To whom reprint orders should be addressed.

0028-0836/82/220395—03$01.00

are recorded on strip-chart (2 mm s^{-1} speed). The number of stations was doubled and a new digital recording system set up in January 1982.

Magnitudes were computed using the relationship $M = 2.\log \tau - 0.9$, where τ is the event duration. The detection limit was $M = -1$ and several events of that magnitude may have been missed. Hypocentres were determined using a homogeneous model with a fixed V_P/V_S ratio of 1.7, and by varying the value of V_P between 4 and 7 km s^{-1}. The inter-station distance was typically 5 km and epicentres located in the network interior are considered accurate to 1 km. For those events which were shallow, errors in depth estimates are ~2 km. For epicentres located outside the network, the errors are larger but <3 km for the events reported here. The corresponding errors on hypocentre depths are probably as high as 5 km. A qualitative check on the locations is provided by the distribution of amplitudes.

The Piton de la Fournaise volcano is characterized by frequent volcanic activity with eruptions following each other at 1–2 yr intervals. Yet the seismicity of the island is remarkably low. Researches on local archives dating from 1800 have revealed no accounts of earthquakes of sufficient magnitude to be felt by the inhabitants outside the immediate vicinity of the volcano. During the 5 months which preceded the eruption, <12 events per month were detected by the network (Fig. 2). Events of magnitude >1 seldom occur and the magnitude 2 level was never exceeded. This is in sharp contrast with the level of background seismicity observed at Kilauea in Hawaii, a volcano which belongs to the same category (intra-oceanic).

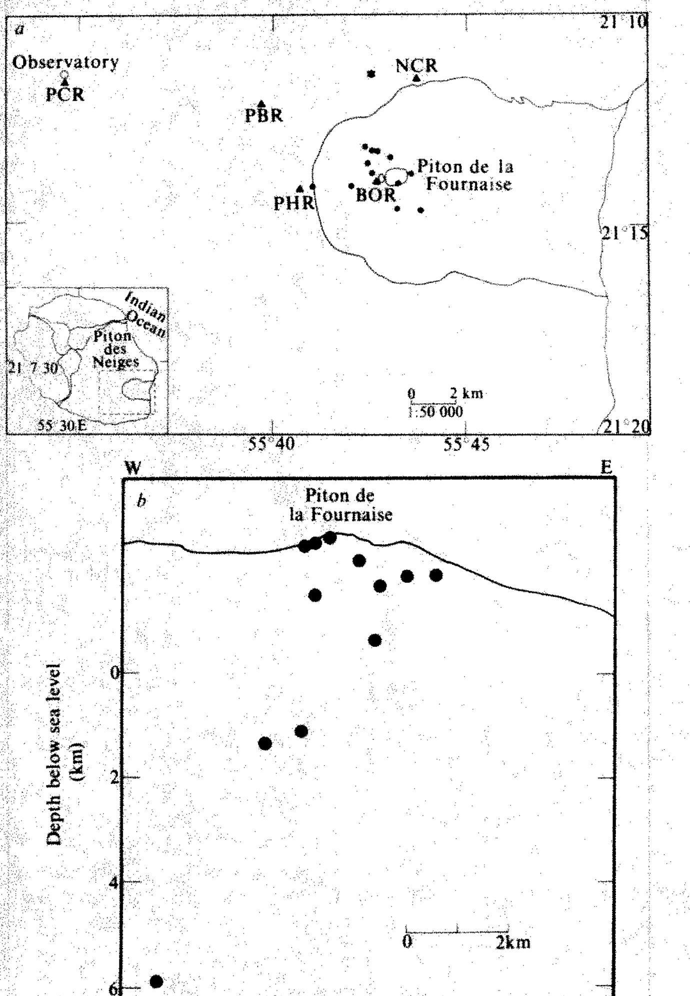


Fig. 1 *a*, Map of the volcano showing the seismic stations (▲) and epicentres of selected events (●). ★, The epicentre of the deep event of 2 January 1981. Locations are accurate to ~1 km in the interior of the network. *b*, Depths of the events shown on *a*. The error is about ±2 km. The deep event at a depth of ~6 km occurred a few hours before the eruption start on 3 February 1981.

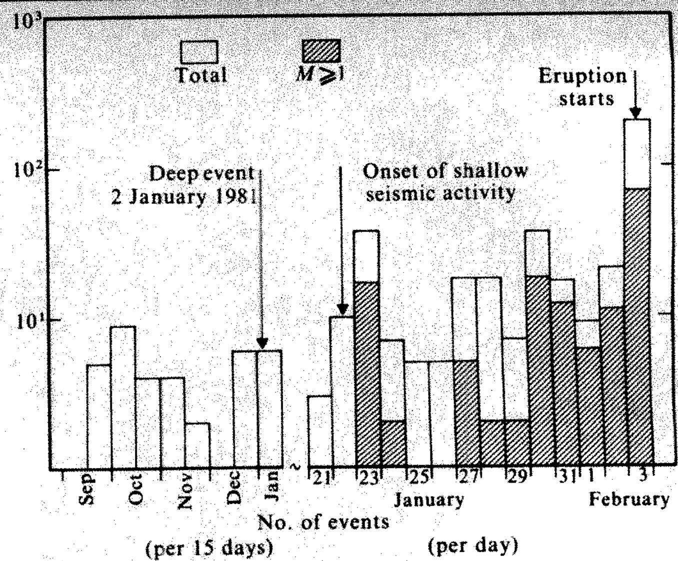


Fig. 2 Seismicity of the Piton de la Fournaise volcano. Note the change of scale on 21 January 1981.

Reports from the Hawaiian Volcano Observatory indicate that the monthly number of events with magnitudes >1 and >3 typically exceed 100 and 10 respectively (see, for example, refs 4, 5).

Premonitory signs of the volcano awakening were first detected on 2 January 1981, when a most unusual event was recorded. It lasted 5 min and was clearly not of tectonic origin (Fig. 3). The S-phase was followed by low frequencies below 0.5 Hz and P-wave polarities varied from station to station. It was a deep event with small-amplitude variations across the network, located 10–15 km below the northern part of the caldera (Fig. 1*a*). It was probably caused by the movement of a light phase (magma or gas) and not by fracturing. No similar event had been observed before and none has occurred since.

No significant increase of seismic activity was noted afterwards (Fig. 2) until 22 January 1981, when 10 events were observed on a single day. On the following day there were 38 events and a warning was issued to the local authorities on 24 January. Seismicity remained at a high level and dramatically increased just before the eruption (Fig. 2). This activity was located around the small Bory crater, close to station BOR (Fig. 1*a*), at depths smaller than 3 km (Fig. 1*b*), with the exception of a deeper event which occurred a few hours before the eruption.

A geodetic network of eight stations was set up around the volcano and the first measurements were carried out on 12 December 1980. The precision in the distance estimates is ~3 cm. Following the onset of seismic activity on 22 January, measurements were conducted on 24 January and 30 January. Only the latter showed a small difference with those of 12 December 1980, barely above the threshold of error, which were interpreted as inflation of the Bory zone[3]. Subsequent measurements made after the eruption start on 12 February 1981 revealed a displacement of 18 ± 2 cm perpendicular to the system of fissures which had opened[3]. No further variations could be detected during February.

Unfortunately no measurements could be made before 12 December 1980. These would have allowed us to see whether

Table 1 Characteristics of the eruptive phases

Eruptive phase	Surface of lava flows (10^6 m^2)	Volume* (10^6 m^3)
1. 3–24 February	0.5	2.5
2. 24 February–31 March	0.7	3.5
3. 1 April–5 May	0.9	4.5

* Computed using a mean thickness of 5 m.

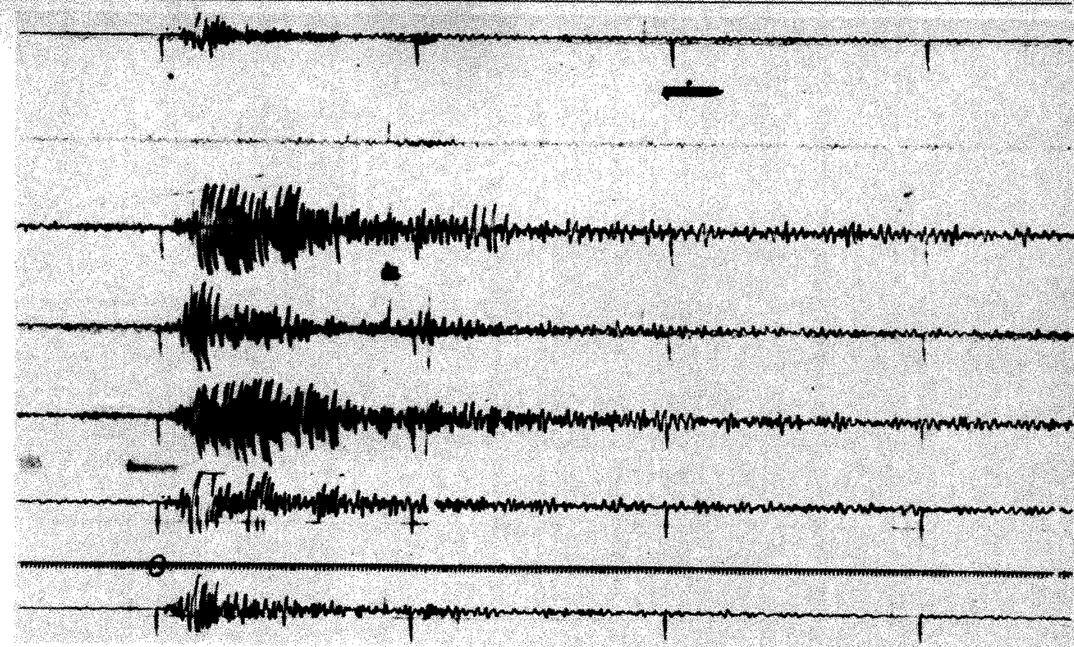


Fig. 3 The 2 Januar event. Large ticks records mark 1 min vals. This event was l at a depth of 10– (from ref. 3).

the inflation detected on 30 January had been going on for a longer time, before the deep volcanic event of 2 January. If such an inflation did occur, however, it was not associated with any significant seismic activity, which would be contrary to what is usually observed at Kilauea[6].

The difference in the seismicity of the two volcanoes may be related to the more complex structure of Kilauea which has several centres of activity. The internal structure of Piton de la Fournaise is not well documented. A detailed study using magnetic data[7] suggests the existence of a small kilometre-sized magma chamber ~1 km beneath the summit area. Shallow seismicity accompanied by inflation is presumably related to the extension of the central dyke complex located immediately below the summit[7].

Eruptive activity occurred in three phases roughly 1 month long (Table 1). Each phase conformed to the same pattern: volcanic tremors for 3–4 h at the start, and virtually no seismicity during the rest of the phase when the greatest volume of lava is erupted. This indicates that tremor activity is due to the initiation of flow.

Tholeiitic lavas came out of fissures which opened along well known NE–SW and NW–SE axes of weakness. They were essentially aphyric[8], contrary to the previous eruptions of April 1977 and July 1979. The three eruptive phases were similar, and characterized by intense degassing with lava fountains reaching heights of 70–100 m. After a couple of days, the activity became limited to a few fissures where spatter cones grew. One or two days later, only one fissure was productive and a volcanic cone was built there. The velocity of lava flow was estimated visually to be close to 4 m s^{-1} at the vent.

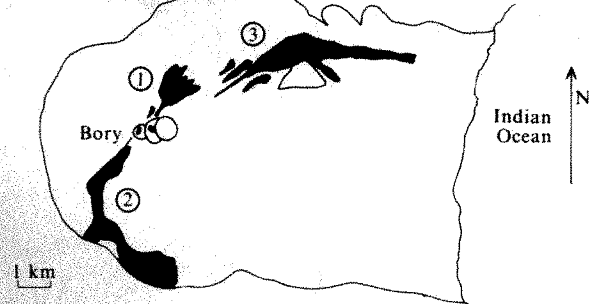


Fig. 4 Map of the lava flows produced by the three eruptive phases.

Lavas produced during this eruption cover large areas (Table 1, Fig. 4) and the total volume is ~10×10^6 m^3, using a mean flow thickness of 5 m. Lava flows were essentially of the aa, gratons and echaudé types which are typical of the island, with some pahoehoe late in each eruptive phase.

Several phreatomagmatic explosions were observed during the first phase, and many during the third, in relation with heavy rainfall.

This is IPGP contribution no. 589.

Received 18 January; accepted 5 April 1982.

1. Wadge, G. *Nature* **288**, 253–255 (1980).
2. Bachelery, P. thesis, Univ. Clermont Ferrand (1981).
3. Blum, P. A., Gaulon, R., Lalanne, F. X. & Ruegg, J. C. *C.R. hebd. Séanc. Acad. Sci., Paris* **292**, 1449–1455 (1981).
4. Koyanagi, R. Y., Meagher, K., Klein, F. W. & Puniwai, G. S. *Hawaiian Volc. Obs. Summary* (1976).
5. Nakata, J. S., Tanigawa, W. R. & Klein, F. W. *Hawaiian Volc. Obs. Summary* (1979).
6. Koyanagi, R. Y., Endo, E. T. & Ward, P. L. *Geophys. Monogr.* **19**, 169–173 (1976).
7. Lenat, J. F. & Aubert, M. *J. Volcan. geotherm. Res.* (in the press).
8. Chevallier, L., Lalanne, F. X., Bachelery, P. & Vincent, P. M. *C.R. hebd. Séanc. Sci. Paris* **293**, 187–190 (1981).

Prehistoric soil and vegetation development on Bodmin Moor, southwestern England

E. Maltby & C. J. Caseldine

Geography Department, University of Exeter, Exeter, Devon EX4 4RJ, UK

Archaeological excavation of barrows in the St Neot Valley on the southern flanks of Bodmin Moor revealed buried soils developed in the gravelly granitic head typical of the upland block. By comparing these soil profiles with adjacent soils we have obtained evidence for and now describe dramatic prehistoric changes in upland soils and vegetation attributed previously to various combinations of human influences and climatic events. Direct pedological and palaeoecological evidence supports the theory that brown soils existed in pre-Bronze Age times in southwestern England at a site subsequently developing an iron pan (Bf) and accumulating a strongly acid peaty surface.

The barrows CRII and CRIV revealed by the excavation have yielded statistically inseparable radiocarbon dates of 1550–1650 and 1560–1630 b.c. (ref. 1) respectively and occupy opposing valley shoulder positions <500 m apart (Fig. 1). The CRIV complex stands in acid grass–heath moorland characteristic of the uplands of south-west England and in which *Molinia caerulea* and *Calluna vulgaris* are locally prominent. The

0028-0836/82/220397—04$01.00

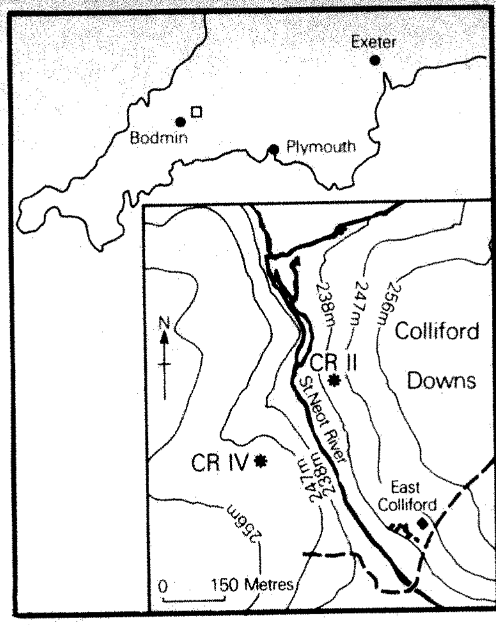


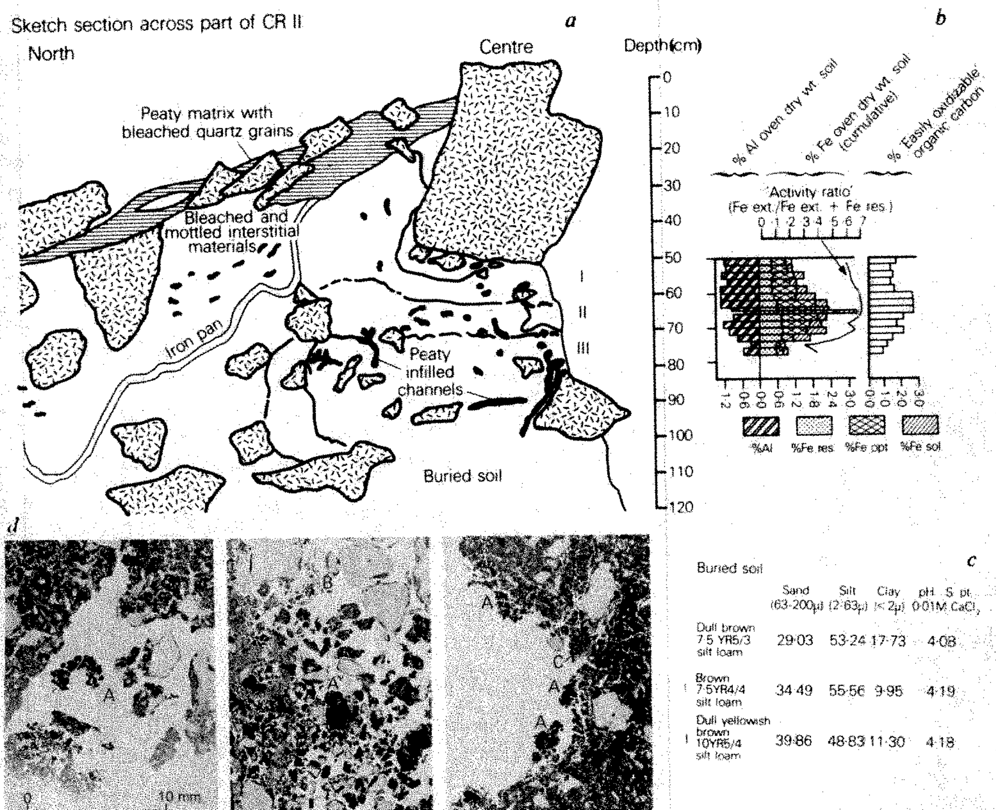
Fig. 1 Location of Bronze Age barrows in the St Neot Valley, Bodmin Moor.

present day soil is a distinctive iron pan stagnopodzol comprising 15–20 cm peaty surface, a variable but stony Ah, grey mottled or pallid E(g) above a thin but wavy and indurated iron pan (Bf) bounding ochreous then brown B_s horizons which merge at 50–60 cm into paler B/C material. CRII occupies a slope enclosed and converted to improved pasture at least since the seventeenth century. Associated soils have a well-mineralized organic surface, commonly lack an E horizon (which has been lost probably by cultivation), but still possess an iron pan (occasionally discontinuous) and strongly developed B_s horizon. It is a cultivated variant of the iron pan stagnopodzol which has been variously mapped and described elsewhere as the Hexworthy Series[2,3].

The profiles beneath two barrow structures of the CRIV complex (CRIVA and CRIVC) are identical to the present day soil except that the original surface peat is attenuated at the old land surface. This is explained by the incorporation of peaty turves into the structure of the barrows. However, a strongly contrasting relationship emerges from the sub-CRII soil. The mineralogenic profile comprises three morphologically distinct horizons above B/C and C(x) material (Fig. 2*a*). Unlike the profile developed within and at the margins of the barrow, under CRIV and associated with the present day surface, there is no organic surface, bleached E_a or E_g horizon or morphological evidence for an iron pan. It seems possible that an original turf layer was removed in the course of barrow construction but otherwise the profile is intact. Horizon notation for the buried profile follows Avery[4] but non-pedogenic nomenclature is preferred for our discussion in view of the potential difficulties in interpretation of buried horizons.

Relatively high values for sodium dithionite extractable iron (Fe_{res}) which are remarkably consistent with depth suggests a strong affinity[5], if not genetic link, between the buried soil and a brown earth type progenitor (Fig. 2*b*). Such crystalline 'aged' Fe only remains in the absence of active biochemical agencies of remineralization usually associated with acid organic surfaces[5] and/or leaf leachates[6,7]. Thus alternatively the pattern of Fe_{res} may be an artefact of burial and protection from fresh sources of complexing or chelating radicals. However, alkaline potassium pyrophosphate extracts (Fe_{ppt}, Fe_{sol}, Al) show particular enrichment of 'active' amorphous Fe in horizon II and confirms general podzolization of the profile. Iron which can be precipitated by NH_3 (Fe_{ppt}) is the dominant fraction in all but the lowest levels sampled. That fraction of the extract remaining soluble (Fe_{sol}) is described by Bascomb as the most active form of Fe linked in 'fresh' organic complexes[5]. Amounts are generally low. This might be expected if the profile were protected substantially from contemporary leaching under the large granite blocks of the barrow but is also a typical feature of brown soils. The proportion of 'active' Fe ($Fe_{ppt}+Fe_{sol}=Fe_{ext}$) to total Fe ($Fe_{ext}+Fe_{res}$) changes with depth in a pattern confirming active enrichment of II and a change to mainly 'aged' iron in underlying B/C material.

A marked peak in active Fe is associated with organic rich infill of a distinct system of channels. The channels are unusual for several reasons: they contain black peaty fill including bleached quartz grains similar to the modern moorland and



	Sand (63-200μ)	Silt (2-63μ)	Clay (< 2μ)	pH S pt. 0·01M $CaCl_2$
Dull brown 7·5 YR5/3 silt loam	29·03	53·24	17·73	4·08
Brown 7·5YR4/4 silt loam	34·49	55·56	9·95	4·19
Dull yellowish brown 10YR5/4 silt loam	39·86	48·83	11·30	4·18

Fig. 2 *a*, Section across part of CRII revealing buried soil especially well preserved beneath large granite block. Note iron pan soil well developed in barrow structure but Bf horizon not extending into buried profile. *b*, Distribution of pyrophosphate extractable Fe, Al and dithionite extractable Fe and 'easily oxidizable' organic carbon (rapid dichromate digestion) in buried profile. *c*, Colour (Munsell notation), textural characteristics (wet sieving and sedigraph analysis), and *p*H ('sticky point' method on fresh material) of buried horizons. *d*, Thin sections across channel features. A, probable earthworm cast material (highly organic); B, root epidermis enclosing small animal droppings; C, void with fine organic detrital material.

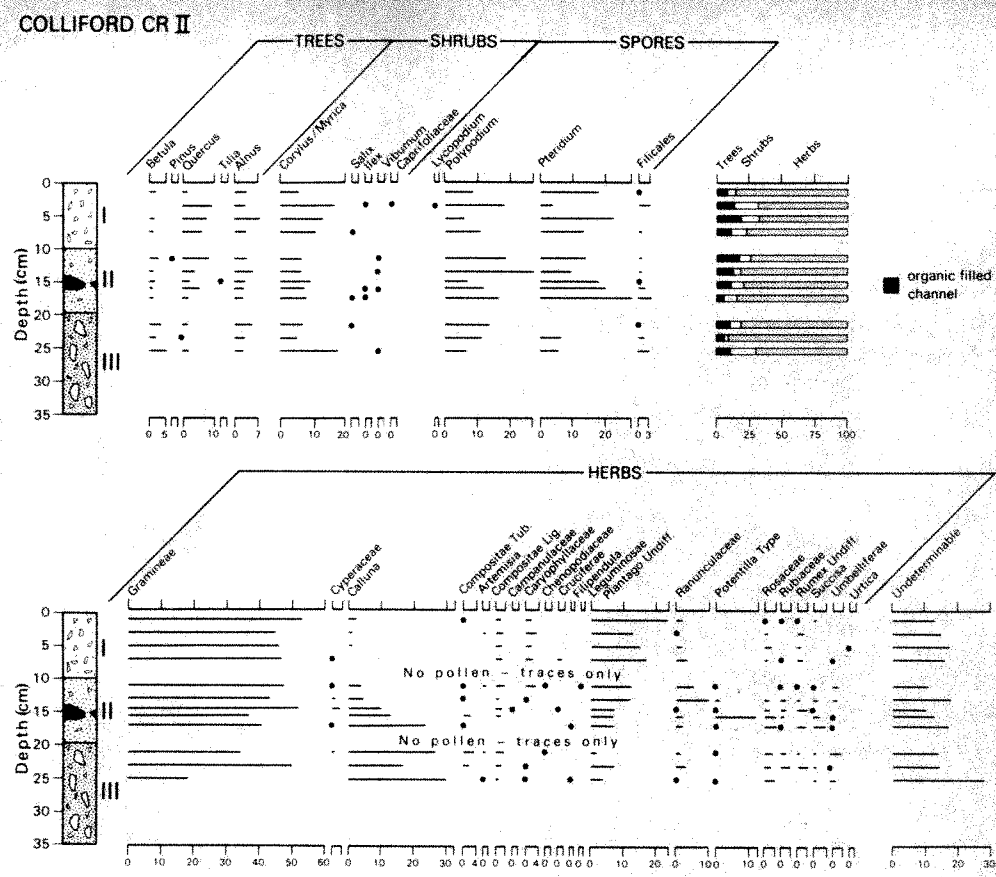


Fig. 3 Pollen diagram from buried soil buried at CRII. All taxa are expressed as a percentage of the total land pollen sum. Percentages for the organic pocket are inserted at the relevent stratigraphic position.

Table 1 Dominant and characteristic pollen taxa from the soil horizons of CRII

Horizon	Pollen taxa
I	Dominated by Gramineae and *Plantago* with *Quercus, Alnus, Betula* and *Corylus/Myrica* comprising 30% T.L.P. Variable but high frequencies for *Polypodium* and *Pteridium*.
II	Dominated also by Gramineae and *Plantago* at reduced levels with reduced *Quercus* and *Corylus/Myrica*. Increasing *Calluna* and greater species diversity including Ranunculaceae, *Potentilla* type, Rosaceae, Rubiaceae, *Succisa* and *Rumex*. Highest values for *Polypodium* and *Pteridium*.
III	Dominated by lower values for Gramineae, and *Calluna*. *Plantago*, *Pteridium*, *Polypodium* and all A. P. are reduced. At the lowest sample there is a peak for *Corylus/Myrica*.
Organic pocket	Similar assemblage to surrounding horizon with dominance of Gramineae, *Plantago* and *Calluna* with *Corylus/Myrica*, *Alnus* and *Quercus*.

barrow surface; they exhibit remarkably little variation in internal diameter (mean = 9.84 mm, $s = \pm 1.12$ mm, $v = 12\%$, $n = 25$) except for occasional enlarged pockets; where channels split or bifurcate they often do so in an upward direction; the internal walls are commonly smoothed; they lack smaller lateral systems and macro-remains of recognizable plant debris such as bracken rhizomes. Thin sections reveal the presence of possible earthworm casts within and lining the channels together with evidence in pores of substantial small animal activity (Fig. 2*d*). We propose that these channels are 'fossilized' burrows of lumbricid earthworms with which they compare in permanence, size and depth[8]. Absence of a highly acidic peat or mor humus surface is necessarily implied by their establishment. The population persisted until soil conditions degraded to a point which allowed accumulation of an acid peaty surface as indicated by the character of the fill in the abandoned burrows. Earthworm activity could be responsible for the much finer texture of horizon I (A/Ah?) and also the general homogeneity of the pollen record within this horizon can be explained by active faunal mixing (Fig. 2*c*). Pollen in the channels themselves is better preserved in the more acid and recent fill than generally in the buried profile. An assemblage with up to 15% *Calluna* confirms the idea of infilling occurring when a peaty surface had begun to develop in association with acidophilous heathland communities. Large earthworms died out at least a short time after 1560 b.c. because several channels extend into the barrow structure. Note in this context that at the experimental earthwork on Overton Down, constructed in 1960, earthworm activity rapidly extended well up into the mound from the buried soil surface. By 1964 worm casts are recorded in interstitial spaces throughout the experimental bank[9].

A pollen diagram has been constructed for the buried soil despite problems caused by generally poor pollen preservation (Fig. 3). Changes in the gross character of pollen assemblages between the three horizons are summarized in Table 1. Pollen evidence suggests that the immediate pre-barrow soil developed in a relatively open grassland environment. Increasing *Calluna* down the profile, however, indicates that heather was an important component of the local vegetation before grassland development. It may have comprised the under-storey or gaps in scrubby woodland dominated by hazel or oak already substantially modified by Neolithic human activity. Indiscriminate burning and/or excessive grazing, together with more cultivation could have led to its demise as pressure for land increased into the Bronze Age.

It has been argued that the Hexworthy Series on Dartmoor has developed from an earlier brown earth type under oak forest[10]. Enhanced acidification and leaching have been associated with Neolithic clearance on land generally above 300 m but because of the temporary nature of such clearings and the inferred frequent regeneration of woodland it is by no means universally agreed that there was any long lasting effect on soil development. Pre-Bronze Age podzolization and iron pan development at ~250–300 m elevation is inferred for Dartmoor and certainly pre-dates barrow construction of this period, including that of CRIV elsewhere on the granite uplands of

southwestern England[13,14]. The present study confirms for the area immediately adjacent to CRII, however, that soil changes leading to iron pan formation and accumulation of a peaty surface post-dates this period. Nevertheless, podzolization was already well enough advanced to produce a brown (possibly humic) podzol profile. A wide variety of buried soils ranging from 'unleached' profiles to thin iron pan podzols were described in association with Bronze Age barrows elsewhere in upland Britain and differences explained in terms of very rapid soil changes occurring within this period of prehistoric activity[15]. High rates of differential pedogenesis cannot be dismissed entirely on Bodmin Moor. However, soil changes are not necessarily time-synchronous even over short distances and the effects of slope aspect, subtle differences in elevation, gradient and hydrology together with any differential influences of prehistoric land management should not be ignored as causes of significant time delays in the processes of conversion of brown soils to iron pan stagnopodzols.

Frances Griffith has coordinated the archaeological programme of which this study has been part. We thank Nigel Fenn and Anne Alderton, financed through the MSC Job Creation Programme, for research assistance and Dr Peter Bullock for preparing thin sections and for his helpful interpretation. South West Water are thanked for site access. Figures are by the drawing office and photographic unit, Geography Department, Exeter University.

Received 1 February; accepted 7 April 1982.

1. Griffith, F. *Corn. Archaeol.* **18,** 56 (1979).
2. Clayden, B. *Soils of the Exeter District* (Soil Survey England and Wales, Harpenden, 1971).
3. Staines, S. J. *Soil in Cornwall I* (Soil Survey Record No. 34, Harpenden, 1976).
4. Avery, B. W. *Soil Classification for England and Wales* (Higher Categories) (Soil Survey Technical Monograph 14, Harpenden, 1980).
5. Bascomb, C. L. *J. Soil Sci.* **19,** 251–268 (1968).
6. Bloomfield, C. *Rep. Welsh Soils Discuss. Group* **11,** 112–123 (1970).
7. Davies, R. I. *Rep. Welsh Soils Discuss. Group* **11,** 133–142 (1970).
8. Edwards, C. A. & Lofty, J. R. *Biology of Earthworms* (Chapman & Hall, London, 1972).
9. Jewell, P. A. & Dimbleby, G. W. *Proc. prehist. Soc.* **32,** 313–342 (1966).
10. Staines, S. J. thesis, Univ. Bristol (1973).
11. Simons, I. G. *Proc. prehist. Soc.* **35,** 203–219 (1969).
12. Staines, S. J. *Proceedings Devon archaeol. Soc.* **37,** 21–47 (1979).
13. Staines, S. J. *Cornish Archaeol.* **14,** 66–67 (1975).
14. Brisbane, M. & Clews, S. *Cornish Archaeol.* **18,** 33–56 (1979).
15. Dimbleby, G. W. *The Development of British Heathlands and Their Soils* (Oxford Forestry Mem. No. 23, 1962).

SO_2 pollution reduces the freezing resistance of ryegrass

Alan W. Davison & Ian F. Bailey

Department of Plant Biology, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, UK

Several authors[1-3] have observed that the effects of SO_2 on growth are more severe in winter and Davies[4] has shown that this may be due to low irradiance and the short photoperiod reducing the plants' capacity to detoxify the pollutant. However, in winter plants are also subjected to potentially lethal stresses such as freezing. The important forage grass, *Lolium perenne*, shows only poor resistance to winter temperatures, hence, any effect of pollution on hardiness could be of considerable significance. Here we demonstrate that SO_2 reduces resistance to sub-zero temperatures and that this effect is influenced by mineral nutrition. A reduction in freezing tolerance may contribute to rapid selection of SO_2-resistant genotypes, particularly in newly sown leys.

Three experiments were performed to determine the effects of SO_2 on freezing resistance. In the first, six 1-week-old seedlings of *L. perenne* cv. S23 were planted in each of 90 7-cm plastic pots containing standard garden potting compost and these were fumigated for 3 weeks in wind tunnels. During this phase of the experiment the day length was 18 h, irradiance 32 W m^{-2} and temperature 15 °C. Controls received <15 μg SO_2 per m^3 and treated plants 250 μg per m^3. SO_2 was monitored using a Meloy SA 285 sulphur analyser.

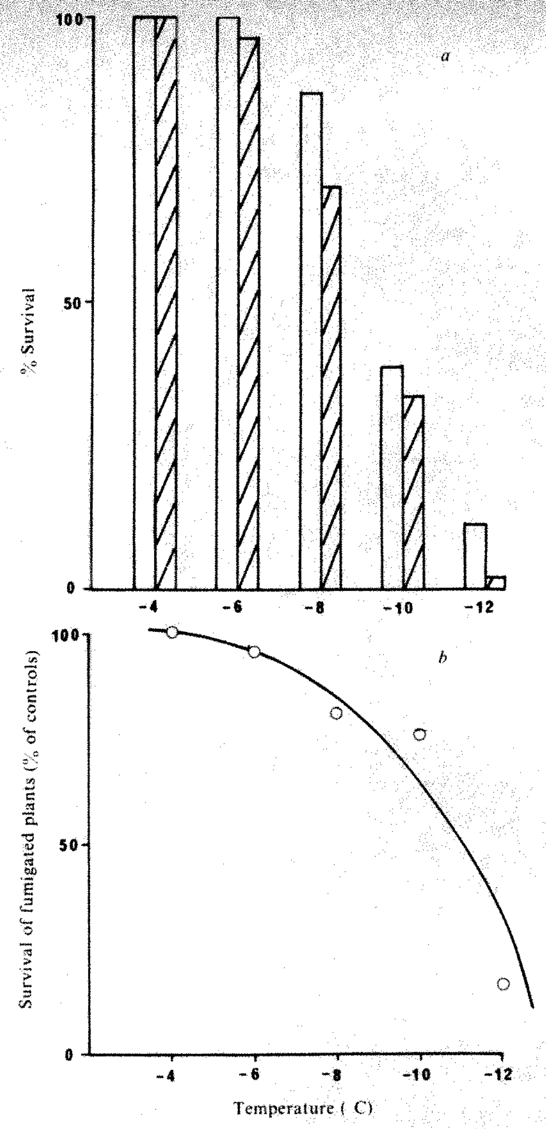


Fig. 1 Survival of *L. perenne* exposed to sub-zero temperatures and SO_2. *a*, Percentage survival of control (open bars, <15 μg SO_2 m^{-3}) and fumigated (cross-hatched bars, 250 μg SO_2 m^{-3}) plants. LT_{50} of controls = -9.7 ± 0.4 °C, that of fumigated plants = -8.8 ± 0.3 °C. *b*, Survival of fumigated plants expressed as a percentage of controls.

After this initial period, fumigation conditions were kept the same except that the temperature was reduced to 2 °C and the daylength to 10 h to 'harden' the plants. After 2 weeks of such treatment, all the plants were removed from the wind tunnels and placed in randomized blocks in a modified deep freeze (daylength, 10 h and irradiance, 55 W m^{-2}). They were not fumigated during this phase. Using a cam-operated temperature control, the plants were held at 0 °C during the first day, followed by exposure during the night to a fall in temperature of 2 °C h^{-1} to a minimum of −4 °C. This minimum was kept constant for 4 h then the temperature was allowed to rise at 2 °C h^{-1} back to the daytime temperature of 0 °C. On subsequent nights the same regime was followed except that the minimum was lowered to give −4, −6, −8, −10 and −12 °C on successive nights. Nine replicate fumigated and control pots were removed at random each day and kept without fumigation at 2 °C in the wind tunnels until all the plants had been removed from the freezer. Then the plants were returned to the same environmental conditions as those during the 3-week pre-hardening period. After 5 weeks of recovery, the plants were scored for survival (post-freezing production of new root, leaf or tiller). The results are shown in Fig. 1.

Survival of the fumigated plants was consistently poorer than controls and probit analysis showed that the LT_{50} values (tem-

Table 1 Influence of soil sulphur and nitrogen on the effects of SO_2 fumigation in *Lolium perenne*

	Low S/ low N	Low S/ high N	High S/ low N	High S/ high N
Control	239*	424†	235*	487‡
($<15\ \mu g\ SO_2\ m^{-3}$)	(22)	(18)	(22)	(16)
$+SO_2$	208*	400†	202*	400†
($250\ \mu g\ SO_2\ m^{-3}$)	(18)	(20)	(12)	(17)

Mean dry wt (mg) of shoots of *L. perenne* cv. S23. Means identified by different symbols are significantly different ($P<0.02$) from each other. Standard errors are shown in parentheses.

peratures at which 50% of the plants were killed) were respectively -8.8 ± 0.3 and -9.7 ± 0.4 °C. Figure 1*b* shows that when survival of fumigated plants was plotted as a percentage of controls the difference between the treatments increased with decreasing temperature.

As the supply of nitrogen and other elements influences[5-9] the effects of SO_2, we performed experiments to determine the effects of soil nitrogen and sulphur on the SO_2–freezing interaction. The first of these experiments was a fumigation–growth study which was used to select contrasting nutrient treatments for use in the second experiment. In the former, newly germinated seedlings of *L. perenne* cv. S23 were sown at a density of 5 per 7-cm plastic pot containing washed, coarse sand. All treatments received deionized water each day and 10 ml of nutrient solution per pot twice a week. Four nutrient treatments were used (Table 1) to give low soil S–low N, low S–high N, high S–low N and high S–high N. All nutrient solutions contained (mg l^{-1}): KH_2PO_4 180, $CaCl_2.2H_2O$ 367, $MgCl.6H_2O$ 313 and trace elements, together with different concentrations of Na_2SO_4 and NH_4NO_3. The low and high sulphur solutions contained 21.3 and 213 mg l^{-1} Na_2SO_4, giving 4.8 and 48 mg S l^{-1}, respectively, while the low and high nitrogen solutions contained 40 and 400 mg l^{-1} NH_4NO_3, giving 14 and 140 mg N l^{-1}, respectively. Twelve replicate pots were used for

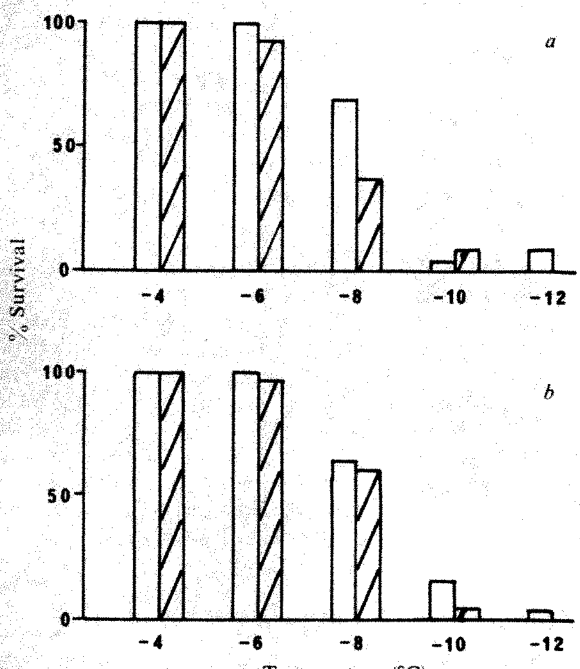


Fig. 2 Survival of *L. perenne* exposed to sub-zero temperatures and SO_2. *a*, Percentage survival of control (open bars, $<15\ \mu g\ SO_2\ m^{-3}$) and fumigated (cross-hatched bars, $250\ \mu g\ SO_2\ m^{-3}$) plants on high sulphur–high nitrogen medium. LT_{50} of controls = -8.4 ± 0.4 °C, that of fumigated plants = -7.6 ± 0.2 °C. *b*, As for *a*, but plants were given low sulphur–low nitrogen medium. LT_{50} of controls = -8.5 ± 0.4 °C, compared with -8.4 ± 0.3 °C for fumigated plants.

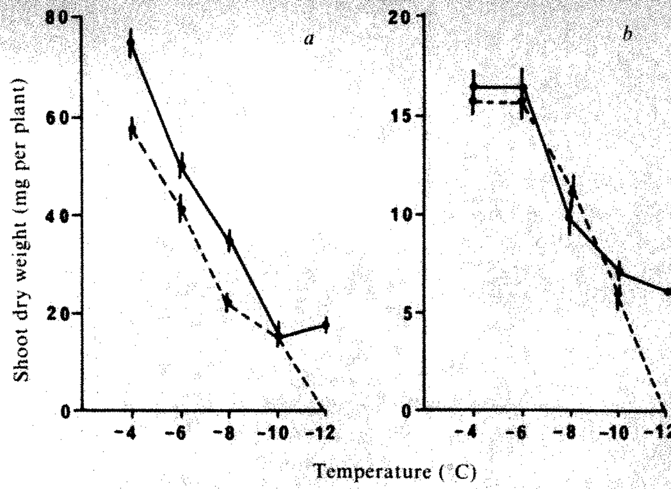


Fig. 3 Dry weight (mg per plant) of *L. perenne* shoots exposed to sub-zero temperatures and SO_2. *a*, Control (—) and fumigated (- - -) plants on high sulphur-high nitrogen medium. *b*, Plants on low sulphur–low nitrogen medium. Vertical bars represent standard errors.

each treatment. Immediately after sowing, the pots were placed in randomized blocks in wind tunnels and exposed to the same environmental conditions as in the pre-hardening period of the first experiment. After 5 weeks of fumigation, the plants were harvested, dried at 105 °C and weighed. The results are shown in Table 1.

In both control and fumigated treatments, an increase in nitrogen concentration significantly increased yield but the effects of soil sulphur were more complex. Increasing the sulphur concentration increased yield only in the presence of high nitrogen concentration in clean air—that is, sulphur was limiting only when the nitrogen concentration was high and SO_2 was low. SO_2 did not appear to compensate for low soil sulphur. With high substrate sulphur and nitrogen, SO_2 caused a significant ($P<0.02$) reduction (18%) in shoot weight. Based on these results, the two extreme nutrient treatments (low S–low N and high S–high N) were used in the final experiment, which examined the interaction of SO_2, substrate sulphur and nitrogen, and freezing injury. Seedlings were grown in sand culture in the same environmental conditions and regime as in the previous experiment. After 5 weeks of recovery from freezing, the plants were scored for survival, harvested, dried and weighed (Figs 2, 3).

The reaction of plants given high substrate sulphur and nitrogen to sub-zero temperatures (Fig. 2*a*) confirmed the results of the first experiment in that SO_2 reduced both survival and the LT_{50} (from -8.4 ± 0.4 °C to -7.6 ± 0.2 °C). However, the difference between control and fumigated plants was greatly reduced (Fig. 2*b*) when they were grown on the low sulphur–low nitrogen medium. The poor winter hardiness of *L. perenne* was reflected not only in the LT_{50} but also in the dry weights. In both nutrient regimes and for both control and fumigated plants, lowering of the temperature caused a progressive decrease in shoot weight but the weight of fumigated plants on the high sulphur–high nitrogen medium was consistently less than controls. On the low sulphur–low nitrogen medium, SO_2 had no effect on shoot weight except at the lowest temperature, where no fumigated plants survived. Root weights showed the same pattern as the shoots (data not shown).

The possibility that pollutants might affect winter hardiness was first mentioned by Bleasdale[10] in 1952 but until now there has been no direct evidence that SO_2 affects any of the attributes that constitute what we call hardiness. A reduction in survival at temperatures that are commonly recorded in British winters has several implications, particularly in relation to efforts to quantify the economic impact of SO_2. Several authors[11-13] have commented on the difficulties involved in extrapolating results obtained in the laboratory to the field situation because of the

effects of variable environmental factors such as light, genetic variation and interactive effects of other pollutants. The finding of an effect of SO_2 on freezing resistance and its modification by mineral nutrition adds to the difficulties and reinforces the view that in order to quantify effects as they occur in the field, experiments must involve exposure of plants at all times of the year and to all the normal range of environmental conditions.

There is good evidence[14-18] that in areas subjected to SO_2 pollution there exists natural selection of tolerant populations. The rate of selection has not been fully established[18] but if SO_2 reduces freezing resistance then this could promote extremely rapid selection, especially in re-seeded leys. It is possible that selection might be so rapid that the effects of SO_2 on the production of re-seeded grasslands might be transient and limited to a short period immediately after sowing.

We thank Alcan (Lynemouth) for financial assistance and Dr W. S. Stewart for the loan of equipment.

Received 25 January; accepted 2 April 1982.

1. Bell, J. N. B. & Clough, W., S. *Nature* **241**, 47–48 (1973).
2. Ashenden, T. W. & Mansfield, T. A. *Nature* **273**, 142–143 (1978).
3. Young, J. E. & Matthews, P. *Envir. Pollut. B* **2**, 353–365 (1981).
4. Davies, T. *Nature* **284**, 483–485 (1980).
5. Cowling, D. W. & Jones, L. H. P. *Soil Sci.* **110**, 346–354 (1970).
6. Leone, I. A. & Brennan, E. J. *Air Pollut. Control Ass.* **22**, 544–547 (1972).
7. Cowling, D. W., Jones, L. H. P. & Lockyer, D. R. *Nature* **243**, 479–480 (1973).
8. Cowling, D. W. & Lockyer, D. R. *J. exp. Bot.* **29**, 257–265 (1978).
9. Ayazloo, M., Bell, J. N. B. & Garsed, S. G. *Envir. Pollut. A* **22**, 295–307 (1980).
10. Bleasdale, J. K. A. *Nature* **169**, 376–377 (1952).
11. Bell, J. N. B. *Nature* **284**, 399–400 (1980).
12. Mansfield, T. A. & Freer-Smith, P. H. *Biol. Rev.* **56**, 343–368 (1981).
13. Davison, A. W. *Br. Ecol. Soc. Symp.*, Essex 1980 (in the press).
14. Taylor, G. E. & Murdy, W. H. *Bot. Gaz.* **136**, 212–215 (1975).
15. Bell, J. N. B. & Mudd, C. H. in *Effects of Air Pollutants on Plants* (ed. Mansfield, T. A.) 87–103 (Cambridge University Press, Cambridge, 1976).
16. Wellburn, A. R., Capron, T. M., Chan, H.-S. & Horsman, D. C. in *Effects of Air Pollutants on Plants* (ed. Mansfield, T. A.) 105–114 (Cambridge University Press, Cambridge, 1976).
17. Horsman, D. A., Roberts, T. M. & Bradshaw, A. D. *Nature* **276**, 493–494 (1978).
18. Bell, J. N. B., Ayazloo, M. & Wilson, G. B. in *Proc. 2nd Eur. Ecological Symp.*, Berlin (in the press).

'*Wasted*', a new mutant of the mouse with abnormalities characteristic of ataxia telangiectasia

L. D. Shultz, H. O. Sweet, M. T. Davisson & D. R. Coman

The Jackson Laboratory, Bar Harbour, Maine 04609, USA

Research into the basic mechanisms underlying the development of complex multi-system diseases has been facilitated by the study of single gene mutations in experimental animals. Ataxia telangiectasia is such a multifaceted genetically determined disease for which no animal model has been described. It occurs in children and is characterized by varying degrees of immunodeficiency, cerebellar ataxia, oculocutaneous telangiectasia, chronic sinopulmonary disease, endocrine abnormalities, chromosomal aberrations, and a high incidence of neoplasms[1-3]. Although the clinical manifestations have been well documented, the cause of the disease is unknown. The pleiotropic nature of the ataxia telangiectasia gene is very complex. Many aspects of this disease, including impairment of tissue differentiation in embryological development are not amenable to direct study, thus an animal model for ataxia telangiectasia would be valuable in elucidating the underlying mechanisms, and in developing effective preventive or therapeutic measures. We have recently found a spontaneous mutation of the mouse called '*wasted*' (*wst*) that shows pathological changes similar to those of ataxia telangiectasia. This mutation can be recognized at 3 weeks of age by neurological abnormalities. We report here that affected animals show pathological changes in the central nervous and lymphoid systems and exhibit a high degree of spontaneous and γ ray-induced chromosomal damage.

The autosomal recessive mutation *wst* occurred spontaneously in 1972 in the inbred HRS/J colony maintained in the Mouse Mutants Stocks Center at the Jackson Laboratory[4]. Homozygotes (*wst*/*wst*) can be recognized at 20 days of age by neurological abnormalities, including tremor and uncoordinated body movements. Affected mice develop progressive paralysis and do not survive beyond 30 days of age. As wasted mice are short-lived, the mutation was crossed to a segregating background (C3HeB/FeJ × C57BL/6J) to increase viability through 'hybrid vigor'. We then studied the effect of the *wasted* mutation on the segregating hybrid background. We are presently backcrossing the *wasted* mutation to the C57BL/6J strain. Although *wst* has not yet been mapped, it does segregate independently of markers on chromosomes 1–11, 13–15, 17 and 19; further linkage tests are in progress. Allelism tests using *ax^J*, *bc*, *cri*, *dw*, *dy*, *mdf*, *med^J*, *myd*, *nr*, *pcd*, *tn*, *twi* and *wl* were all negative. (For a description of these genes, see ref. 5.)

While mice normally show rapid growth and weight gain between 20 and 30 days of age, *wasted* homozygotes show no weight increase after 20 days. Of particular interest was the finding that, by 28 days of age, *wasted* mice have a marked lymphoid hypoplasia with significantly decreased ($P<0.05$, Student's *t*-test) spleen, thymus, and lymph node to body weight ratios (Table 1). Analysis of peripheral blood in 17 pairs of mutant and control mice revealed that 3–4-week-old wasted homozygotes have a threefold decrease in numbers of circulating leukocytes ($2.8\pm0.52\times10^6\ \mathrm{ml}^{-1}$ for *wst*/*wst* mice as opposed to $8.7\pm1.06\times10^6\ \mathrm{ml}^{-1}$ for +/− littermates). To determine whether the *wst* mutation affects development of both Thy-1-bearing cells (T cells) and surface immunoglobulin-bearing cells (B cells), spleen cell suspensions were analysed by indirect immunofluorescence microscopy[6] using monoclonal anti-Thy 1.2 antibody (cell culture supernatant from hybridoma HO-13-4.9)[7] and fluorescein-conjugated goat anti-mouse immunoglobulin. Preliminary analysis of nine pairs of wasted and littermate control mice (20–28 days old) showed no significant effect of the *wst* mutation on percentages of T and B cells (data not shown). Thus, the lymphoid hypoplasia is not restricted to a single major lymphocyte subpopulation. Similarly, in ataxia telangiectasia patients, percentages of circulating B cells are normal while those of T cells are within the lower normal limits[8].

As patients having ataxia telangiectasia show impairment of delayed-type hypersensitivity (DTH), we examined the DTH response of wasted mice to sheep red blood cells (SRBC). Homozygotes (*wst*/*wst*) had significantly decreased DTH responses 4 days after priming compared with +/− littermate controls, as measured by the specific increase in footpad thickness 24 h after injection of an eliciting dose (10^8) of SRBC[9] (data not shown).

Histological examination of tissues from 10 pairs of *wasted* and littermate control mice (22–29 days old) revealed abnormalities in the central nervous and lymphoid systems of mutant mice. The brain showed degeneration of Purkinje cells and focal demyelination in the cerebellar cortex (Fig. 1); focal demyelination was also found in the ventral columns of the spinal cord of wasted mice. Purkinje cell degeneration and focal demyelination have also been described in association with ataxia telangiectasia[1,3,10]. Marked hypoplasia was found in both the thymus-dependent and thymus-independent areas of the lymphoid system. The follicles of the lymph node and spleen

Table 1 Body weight and lymphoid organ/body weight ratios of wasted mice

Genotype	Body weight (g)	Lymphoid organ weight (mg)/body weight (g): Thymus	Spleen	Lymph node*
wst/*wst*	7.8±0.51	2.4±0.34	2.0±0.23	0.19±0.03
+/−	14.8±0.61	5.2±0.29	4.3±0.17	0.27±0.02

Values are mean±s.e.m. ($n=8$). Lymphoid organs were taken from 28-day-old wasted and littermate control mice of the same sex.
* Right brachial lymph nodes were weighed.

0028–0836/82/220402—03$01.00

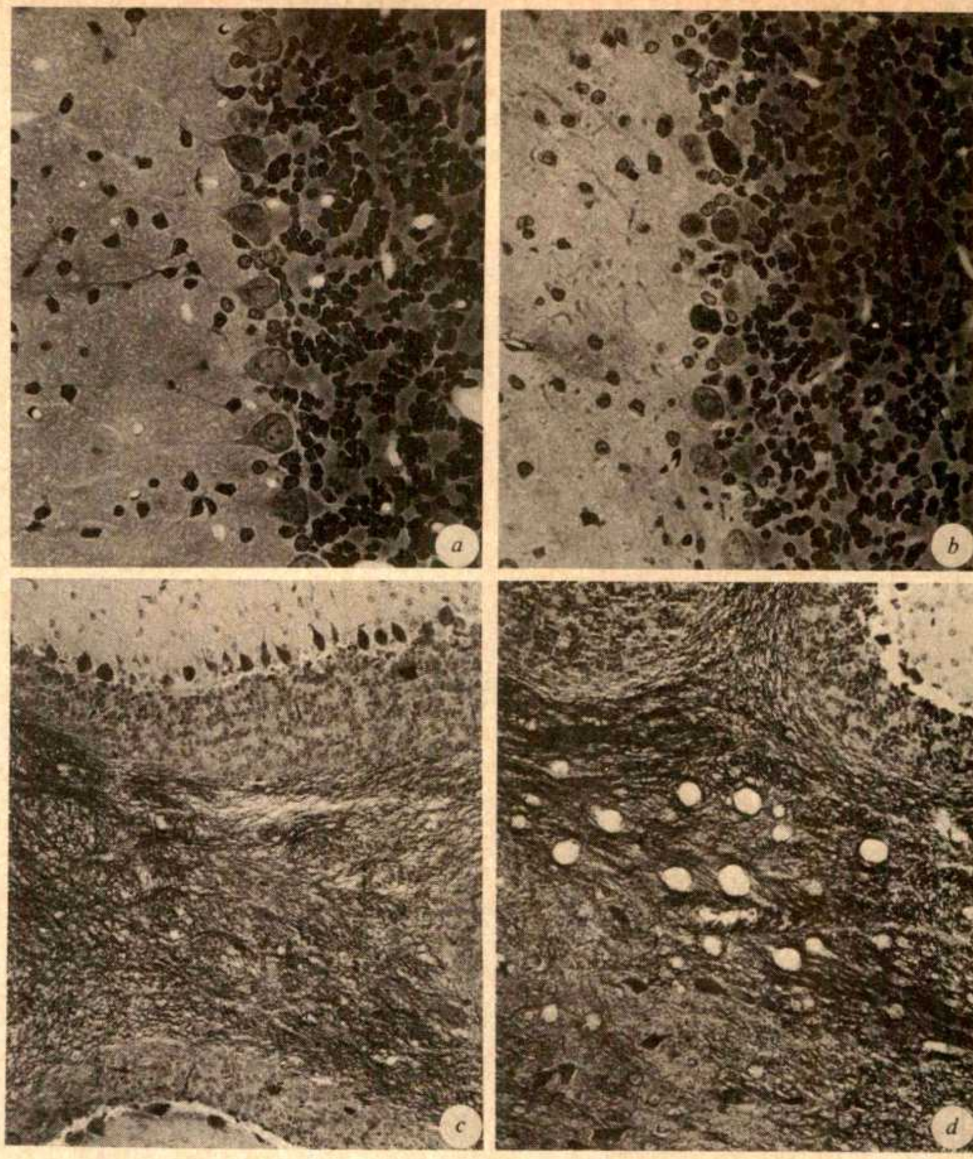


Fig. 1 Sections of brain from 29-day-old wasted and littermate control mice. *a*, Cortex of cerebellum from littermate control showing normal Purkinje cells; aborescent fibrils are evident. *b*, Cortex of cerebellum from *wst/wst* mouse; Purkinje cells appear shrunken and deficient in fibrils (luxol fast blue–cresyl violet, ×140). *c*, Cerebellum from littermate control showing normal myelination. *d*, Cerebellum from *wst/wst* mouse showing advanced myelin loss (luxol fast blue–cresyl violet, ×70).

were poorly developed, Peyer's patches were few and small, and there was a reduction in cellularity of the thymic cortex (Fig. 2). Thymic hypoplasia and lymphocyte depletion in the cortical and paracortical areas of lymph nodes are common features of ataxia telangiectasia[3]. No gross or histological evidence of telangiectases was found in wasted mice, but development of such vascular abnormalities is thought to require exposure to sunlight[11].

As ataxia telangiectasia is characterized by spontaneous and induced chromosomal aberrations, such changes were sought in wasted mice. We examined chromosomes in bone marrow cells collected from nine pairs of *wasted* and littermate control mice (23–28 days old). Cells were not G-banded so that they could be scored for chromosome gaps. At least 50 cells from each mouse were scored blindly for chromosome damage. The *wst/wst* mice showed a fourfold greater incidence of chromosomal damage then their littermate controls ($8.8 \pm 1.69\%$ of cells from *wst/wst* mice compared with $2.2 \pm 0.77\%$ of cells from $+/-$ littermates showed chromosomal damage). The damage included chromosome and chromatid breaks, gaps and fragments, implying either increased chromosome breakage or failure of breaks to rejoin. Preliminary data suggest that wasted mice may also be more susceptible to radiation-induced chromosomal injury. Two *wst/wst* mice and two littermate controls were irradiated with 200R from a Shepard Mark 1 irradiator (^{137}Cs). After 18 h the mice were treated with colchicine, and chromosomes were prepared from bone marrow cell suspensions; 80% of 30 metaphase cells in *wst/wst* mice showed damage of some sort while the controls revealed only minor damage in 30% of the cells scored.

Thus, homozygosity for the wasted mutation results in impaired neurological and immunological development as well as increased levels of spontaneous and γ ray-induced chromosomal damage Whether the chromosomal aberrations in *wasted* mice are associated with defective DNA repair, as has been shown in cell lines isolated from individuals with ataxia telangiectasia[1], has not been determined. Elucidation of the

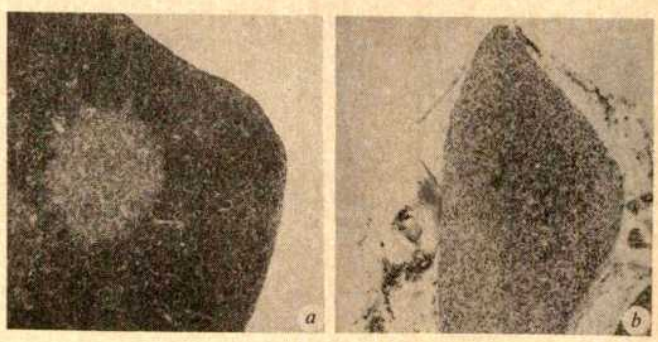


Fig. 2 Thymuses from 29-day-old wasted and littermate control mice. *a*, Littermate control showing normal thymic structure; *b*, 'wasted' thymus showing depletion of cells from cortex (haematoxvlin–eosin, ×25).

mechanism of action of the wasted mutation may further our understanding of ataxia telangiectasia.

We thank Cheryl Bailey for technical assistance. This work was supported by NIH research grants CA20408 and GM27102, NSF grant DEB79-26708, and ACS grant CD-34U.

Note added in proof: Associated with the focal demyelination in the brain and spinal cord of wasted mice are numerous dilated capillaries.

Received 15 March; accepted 13 April 1982.

1. Paterson, M. C. & Smith, P. J. *A. Rev. Genet.* **13**, 291–318 (1979).
2. Kraemer, K. H. in *DNA Repair Processes*, 37–71 (Miami Symp. Spec., 1977).
3. McFarlin, D. E., Strober, W. & Waldmann, T. A. *Medicine* **61**, 281–314 (1972).
4. Sweet, H. O. *Mouse News Lett.* **65**, 27 (1981).
5. Green, M. C. *Genetic Variants and Strains of the Laboratory Mouse* (Fischer, New York, 1981).
6. Sidman, C. L., Shultz, L. D. & Unanue, E. R. *J. Immun.* **121**, 2392–2398 (1978).
7. Marshak-Rothestein, A. *et al. J. Immun.* **122**, 2491–2497 (1979).
8. Griscelli, C. in *Immunodeficiency in Man and Animals*, 45–49 (Sinauer, Stamford, Connecticut, 1975).
9. Miller, T. E., Mackaness, G. B. & Lagrange, P. H. *J. natn. Cancer Inst.* **51**, 1669–1676.
10. Terplan, K. L. & Krauss, R. F. *Neurology* **19**, 446–454 (1969).
11. Reed, W. B., Epstein, W. L., Boder, E. & Sedwick, R. *J. Am. med. Ass.* **195**, 746–753 (1966).

Peripheral pathways are pioneered by an array of central and peripheral neurones in grasshopper embryos

Robert K. Ho & Corey S. Goodman

Department of Biological Sciences, Stanford University, Stanford, California 94305, USA

During embryonic development, neuronal growth cones traverse long distances along stereotyped pathways to arrive at their appropriate targets. Since Harrison's initial observations[1], increasing evidence has supported the notion that axonal pathways are "pioneered" early in development when the distances are short and the terrain relatively simple; thus an axonal scaffold is erected upon which later neurones can navigate. Grasshopper embryos have emerged as an attractive preparation in which to study pathfinding because the nervous system is relatively simple, and the individual identified neurones in the periphery and central nervous system (CNS) of the embryo are large and highly accessible. How are the first axonal pathways established between the periphery and CNS? Bate[2] has described a single pair of cells in the limb buds, and two pairs in the antennae, which were thought to establish the very first peripheral pathways over distances of several hundred micrometres from the periphery to CNS; he called these pairs of cells 'pioneer neurones'. Keshishian[3] confirmed this finding and also described[4] a second pair of pioneers in the limb buds. We report here a further analysis of peripheral pathfinding in the grasshopper embryo using a monoclonal antibody, called I-5, which stains the cells responsible for establishing the first pathways[5,6]. In contrast to the earlier reports, we find a large array of peripheral cells responsible for establishing the first pathways, with both pathfinders and landmark cells spaced at short intervals along the route. Furthermore, growth cones from central motoneurones make important contributions to the first peripheral pathways[6].

Monoclonal antibodies against the grasshopper nervous system have been made with the goal of obtaining cell markers which label specific subsets of cells early in embryogenesis[5,6]. In early embryos, one of these antibodies, called I-5, stains a specific pattern of neurones and mesodermal cells[5,6], including the pairs of peripheral pioneer neurones originally described by Bate[2] and Keshishian[3,4]. Interestingly, the I-5 antibody also reveals an array of previously unseen cells and growth cones that are intimately involved in pioneering the first peripheral axonal pathways. In this paper we have studied these cells using the I-5 monoclonal antibody[5,6], observations with Nomarski optics[7,8], intracellular injections of the fluorescent dye Lucifer Yellow[9], and an anti-Lucifer Yellow antibody (anti-LY)[10].

The I-5 antibody stains the pairs of pioneer neurones in both the antennae[2] and limb buds[2–4] shortly after their final cell division, and before axonogenesis (Fig. 1*a*); the antibody also stains their growth cones and axons (Fig. 1*b*,*c*). In the antenna (Fig. 2), two pairs of neurones [first the ventral pioneers (VP) and then the dorsal pioneers (DP)] appear at the distal tip and extend growth cones towards the CNS, thus pioneering the two axonal pathways within the antenna. At the same time, a single base pioneer (BP) appears at the base of the antenna and its growth cone is the first peripheral process to reach the CNS. Simultaneously, a growth cone from a motoneurone originating in the CNS extends outwards towards the base of the antenna, running parallel, although not fasciculating upon, the BP's axon; the motoneurone growth cone stops at the lateral edge of the antennal base and in time innervates a muscle that develops at this location. Thus, the pathway between the antennal base and the CNS is complete by the time the growth cones from the more distal VP and DP neurones arrive and follow the BP pathway into the CNS. Interestingly, when these cells first appear inside the ectodermal epithelium, they are spaced at

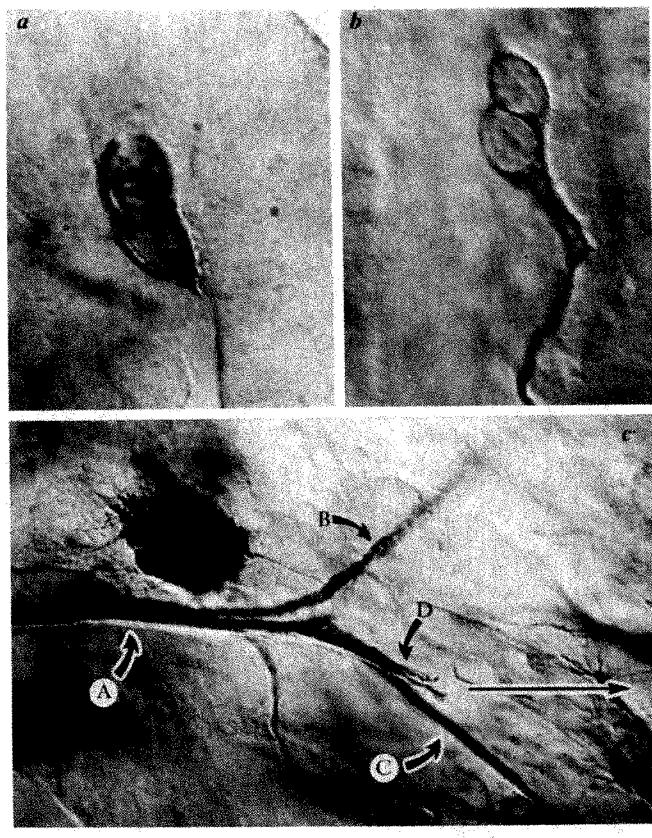


Fig. 1 Photomicrographs of limb buds (A, C) and antennae (B) showing staining of cells with I-5 monoclonal antibody; the growth cones of these cells are involved in pioneering the peripheral axonal pathways in the grasshopper embryo. The embryo is dissected out of the eggshell, lightly fixed in 2% paraformaldehyde, and incubated overnight in I-5 supernatant with 0.2% saponin. The antibody is visualized using an HRP-conjugated second antibody. *a*, Staining of the pair of 1E cells in the limb bud just before they initiate growth cones. The 1E cells are just within the ectodermal epithelium at the tip of the limb bud. *b*, Staining of the ventral pioneers (VP) in the antennae after they have initiated axons. Their growth cones (not shown) have just arrived at the cell body of the base pioneer (BP). *c*, Staining of the limb bud showing the intersection of the A, B, and C pathways (established by the 1A, 1B, and 1C cells, respectively), and the growth cones of the early motoneurones (the β motoneurones) as they grow distally and establish the D pathway (long arrow) (see Figs 3, 4). Large cell above A pathway is a mesodermal cell also stained by the I-5 antibody. Cell body diameters in *a* and *b* are ~15 μm; large mesodermal cell in *c* is ~30 μm in diameter.

0028-0836/82/220404—03$01.00

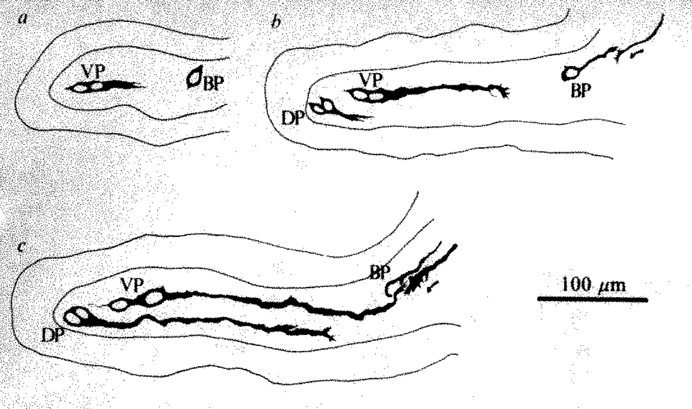


Fig. 2 Drawings of the antennae of the grasshopper embryo at successively older stages of development, showing the pioneering of the peripheral axonal pathways by the growth cones from peripheral and central (arrows) cells. Bar, 100 μm.

relatively short distances from one another; the VP neurones are less than 100 μm from the BP neurone, and the BP neurone is less than 100 μm from the CNS. Lucifer Yellow injections (and visualization with anti-Lucifer Yellow) reveal that the cells before axonogenesis and the growth cones afterwards radiate long, thin filopodia that typically reach over 50 μm and sometimes for up to 100 μm (ref. 10). Thus, the cells may be within filopodial grasp at, or shortly after, axonogenesis.

A similar, albeit more complex, series of events takes place in developing limb buds (Figs 3, 4). The first growth cones in the limb bud arise from the 1B cells (Bate's[2] original 'pioneers'). The 1B growth cones extend proximally towards the CNS along the basement membrane on the inside of the ectodermal epithelium of the limb bud, growing past where the 2B cell appears and towards the 3B cell (the B pathway, Fig. 4). However, the 1B growth cones make a characteristic turn about the time they reach the 3B cell (although they need not reach this cell); they always, however, turn towards the 1A cells which are about 50 μm away and within filopodial grasp (Fig. 3). Upon reaching the 1A cell bodies, the 1B growth cones then continue extending towards the CNS. One source of guidance for the 1B growth cones is likely to be the 1A cells (as 'landmark' cells). In the pro- and mesothoracic limb buds, the 1A cells send growth cones towards the CNS before the 1B growth cones arrive at their cell bodies; thus the 1A cells pioneer the first pathway (the A pathway, Fig. 4) between the base of the limb bud and the CNS (Fig. 3*a*). The timing is different, however, in the metathoracic limb bud. Here the 1A cells do not pioneer the first pathway to the CNS; rather, the 1B cells grow towards them and then continue growing towards the CNS, followed shortly thereafter by the 1A growth cones.

A second pathway is simultaneously made by growth cones of motoneurones originating in the CNS; these growth cones do not fasciculate with the 1A or 1B axons but rather form a separate yet parallel pathway between the base and the CNS. As development proceeds, the A pathway becomes nerve 5 whereas the second parallel pathway becomes nerve 3.

The C pathway is made by the pair of 1C neurones. Much of the D pathway is pioneered by growth cones of motoneurones originating in the CNS which extend out along the A pathway shortly after it is established; instead of growing along the B or C pathways, these motoneurones make a new pathway as they grow distally (Figs 1, 3, 4). These motoneurone growth cones meet those from the 1D cells and thus the D pathway is completed. The E and F pathways are pioneered by pairs of peripheral cells in a similar fashion simultaneous to the pioneering of the D pathway (Figs 3, 4) [the 1E cells are Keshishian's[4] second pair of pioneers]. Just as in the antenna, the distances between the peripheral neurones in the limb bud are relatively short (often 50 to 100 μm) when they first extend growth cones. Here too the cells have extensive filopodia and may be in direct contact at, or shortly after, axonogenesis.

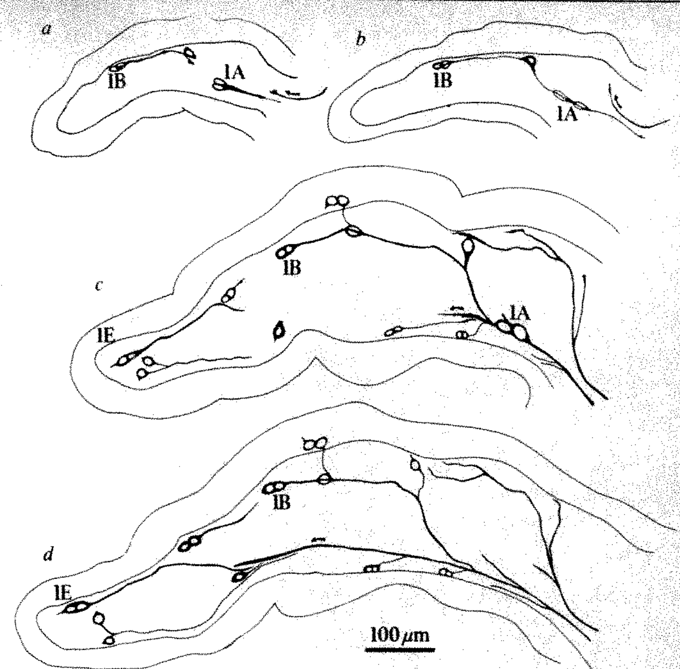


Fig. 3 Drawings of the mesothoracic limb bud of the grasshopper embryo at successively older stages of development, showing the pioneering of the peripheral axonal pathways by the growth cones from peripheral and central (arrows) cells. For clarity, only a few cells are named; the complete identities are shown in Fig. 4.

Four major conclusions are suggested by our results. First, peripheral pathways are pioneered by growth cones from central as well as peripheral neurones. For example, nerve 3 and the D pathway of nerve 5 are pioneered by motoneurones.

Second, there are not just two pairs of cells involved in pioneering the peripheral pathways in limb buds, but rather a large array of cells. In sheer number, this array of cells (Fig. 4) at first glance appears to hopelessly increase the complexity of the problem. Yet their unveiling dramatically simplifies the picture and points us towards specific experiments. The finding that the cells are so close and no one cell must alone traverse a large expanse of axonless territory, shifts our attention from long distance guidance mechanisms (e.g. diffusible gradients) to local interactions between identified cells over short distances. In many cases, cells may be in direct filopodial contact with 'landmark' cells (for example the 1A cells appear to be stepping stones [2] for the 1B growth cones); in other cases, specific mechanical substrates (e.g. the tendon followed by the 1E cells, see below) may take the growth cones to their next neuronal contact. The finding of landmark cells in the periphery is similar to our finding of landmark cells in the CNS[12].

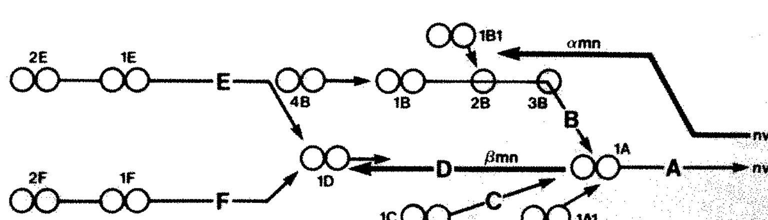


Fig. 4 Schematic diagram showing the identities of cells responsible for pioneering the pathways A to F of nerve 5, and the first pathway of nerve 3 is the pro- and mesothoracic limb buds. In the metathoracic limb bud, the 1A cells do not pioneer the A pathway, but rather the 1B growth cones use the 1A cells as landmarks and then continue growing towards the CNS.

Third, growth cones in the periphery are not simply guided by passive mechanisms. Active guidance mechanisms are suggested because certain growth cones are confronted with several possible choices and make cell-specific decisions of which way to go. For example, the 1E growth cones extend along the epithelium until they near the 4B cell bodies. Rather than climb onto the B pathway, they make a turn and grow along a tendon to the 1D cells, thus completing the E pathway. In a similar way, early motoneurone growth cones are confronted with the B and C pathways, yet choose to grow laterally and pioneer a new D pathway. The finding of specific choices made by growth cones in the periphery is similar to the divergent and specific choices made by identified growth cones in the CNS[11-14].

Finally, we must re-evaluate the concept of 'pioneer neurones'. Bate and Goodman have shown that other peripheral nerves (such as the intersegmental nerve and the median nerve) are pioneered by motoneurone growth cones from neuroblast progeny in the CNS (manuscript in preparation). Furthermore, pathways in the CNS are pioneered by progeny from two different classes of neuronal precursors—the midline precursors and neuroblasts[12,15]. Thus, there is no specialized class of precursor cell which gives rise to cells having the unique role of establishing axonal pathways throughout the grasshopper embryo. Rather, neurones of diverse ectodermal origin appear early in embryogenesis and pioneer pathways when the distances are short and the terrain relatively simple. This finding shifts the emphasis away from the special role played by 'pioneers', and focuses our attention on the questions of what cell-to-cell interactions and substrate features guide the first peripheral growth cones. These questions can be answered by manipulating the pioneer neurones, landmark cells, and ectodermal epithelium in grasshopper embryo cultures. In this way, we hope to establish by what mechanisms these 'pioneering' growth cones are guided across short distances of axon-less territory.

We thank Susannah Chang for the I-5 antibody and Paul Taghert and Michael Bastiani for the anti-LY antibody. This work was supported by grants from the McKnight Foundation, NSF, and Sloan Foundation to C.S.G.

Received 30 November 1981; accepted 2 April 1982.

1. Harrison, R. G. *J. exp. Zool.* **9**, 787–846 (1910).
2. Bate, C. M. *Nature* **260**, 54–56 (1976).
3. Keshishian, H. *Devl Biol.* **80**, 388–397 (1980).
4. Keshishian, H. & Bentley, D. *Soc. Neurosci.* **7**, 347 (1981).
5. Chang, S., Ho, R., Raper, J. A. & Goodman, C. S. *Soc. Neurosci.* **7**, 347 (1981).
6. Ho, R., Chang, S. & Goodman, C. S. *Soc. Neurosci.* **7**, 348 (1981).
7. Goodman, C. S. & Spitzer, N. C. *Nature* **280**, 208–214 (1979).
8. Goodman, C. S., O'Shea, M., McCaman, R. E. & Spitzer, N. C. *Science* **204**, 1219–1222 (1979).
9. Stewart, W. W. *Cell* **14**, 741–759 (1978).
10. Taghert, P., Bastiani, M., Ho, R. K. & Goodman, C. S. *Cell* (submitted).
11. Raper, J. A. & Goodman, C. S. *Soc. Neurosci.* **7**, 347 (1981).
12. Goodman, C. S., Raper, J. A., Ho, R. & Chang, S. *40th Symp. Soc. dev. Biol.* (in the press).
13. Raper, J. A., Bastiani, M. & Goodman, C. S. *J. Neurosci.* (in the press).
14. Raper, J. A., Bastiani, M. & Goodman, C. S. *J. Neurosci.* (in the press).
15. Bate, C. M. & Grunewald, E. B. *J. Embryol. Exp. Morphol.* **61**, 317–330 (1981).

Electrophysiology of mammalian thalamic neurones *in vitro*

Rodolfo Llinás & Henrik Jahnsen

Department of Physiology and Biophysics,
New York University Medical Center,
550 First Avenue, New York, New York 10016, USA

Although much is known about the morphology and physiology of thalamic neurones[1], no information is available regarding the ionic basis for the electrical excitability of these cells. Furthermore, possible differences in the electrical properties of the principal nerve cells in the various thalamic groups have not been studied in sufficient detail to determine whether the thalamus is, electrophysiologically, a uniform set of neuronal elements. Here we present evidence that thalamic neurones have voltage-sensitive ionic conductances capable of generating two distinct functional states—a repetitive and a burst-firing mode. The neurones are switched from one state to the other by membrane potential changes, each state being dominated by different voltage-dependent ionic conductances. At membrane potentials more positive than −60 mV, the neurones respond to a depolarization with repetitive firing via Na^+-dependent action potentials, whereas at potentials more negative than −65 mV, depolarization of the cell results in short bursts of action potentials via an inactivating all-or-none Ca^{2+}-dependent spike. This property, present in all the neurones comprising the different thalamic nuclei, serves as the basis for their oscillatory properties. Particularly, the inactivating Ca^{2+} conductance represents the ionic basis for the post-anodal exaltation, the mechanism most probably responsible for the alpha rhythm.

The study was performed on 400-μm-thick thalamic slices obtained by coronal section of guinea pig brain using a vibratome cutting device (Oxford G501 Vibratome). Typically, six sections included most of the thalamus. The sections were placed in a special lucite chamber under a flowing mammalian Ringer's solution[2]. Electrophysiological analysis was performed, under direct microscopic observation, on neurones of the anterior, medial and lateral thalamic nuclear groups, including the lateral geniculate nucleus and neurones in the lamina interna. The techniques for sectioning and maintenance of the slice have been described elsewhere[2]. Intracellular recordings were made using potassium citrate-filled micropipettes having a d.c. resistance of 50–90 MΩ. Cells were activated either antidromically or synaptically by means of bipolar stimulating electrodes located on the midline region of the thalamic mass (see arrow in Fig. 1*A*) or directly through the microelectrode using a bridge amplifier. Neurones were identified by their antidromic and synaptic activation, and their thalamic site was determined by direct observation. Although single somata were often not visible, the exact outline of the groups of nuclei could be easily determined (see Fig. 1*A*). Intracellular recordings from these neurones indicated that they could survive well for >8 h after the slicing procedure. The stimulus amplitude required for the antidromic activation of these cells (Fig. 1*B*) was often lower than that for the activation of synaptic inputs (Figs 1*C*, 2*E*, *F*). Considering the thickness of the slices, we suspect that only monosynaptic connectivity can be reliably studied in this material.

The electrical properties of the cells were examined by direct stimulation via the recording electrode. As shown in Fig. 2*A*, thalamic cells may be characterized by two distinct types of electrical response. When the resting potential was more positive than −60 mV, thalamic neurones fired repetitively at increasing frequencies with increasing amplitude of direct depolarization. A plot of the instantaneous adapted firing frequency, f (1 s pulse duration) against current injection (I) is shown in Fig. 2*C*, which illustrates the graded nature of the repetitive firing properties of these cells. However, in contrast to other central neurones[3], no primary firing range was observed: rather, the neurones fired with a typical minimal frequency of ~10 impulses s^{-1} and reached firing frequencies of 165 impulses s^{-1}. At this membrane potential level, we observed no after-depolarization comparable to that seen in motoneurones[4], and Purkinje[2], inferior olivary[5] and hippocampal[6] cells *in vitro*. On the other hand, when the cells had resting potentials more negative than −60 mV, or were hyperpolarized by direct inward current injections, rather than the continuous firing, a single burst of action potentials was observed after a step depolarization (see Fig. 2*A*, *B*). The lower record in Fig. 2*A* demonstrates a passive membrane depolarization produced by a square current pulse that is subthreshold for spike initiation. When, as shown in the upper record in Fig. 2*A*, the same

0028-0836/82/220406—03$01.00

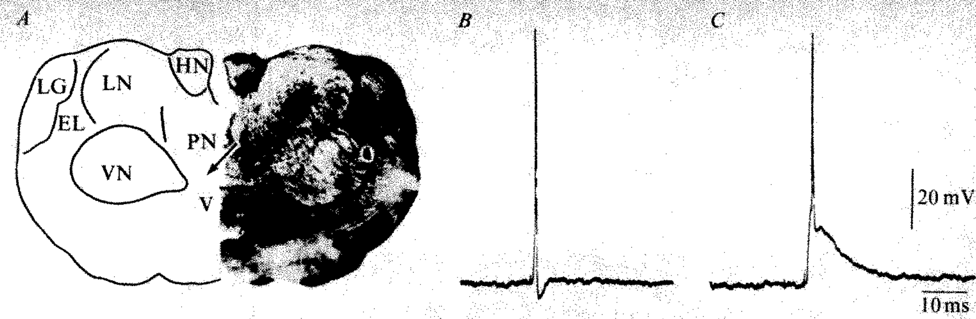


Fig. 1 Neuronal localization in thalamic nuclei. *A*, photomicrograph of diencephalic slice (right) and diagram illustrating the location of thalamic nuclei (left). EL, external medullary lamina; HN, habenular nucleus; LG, lateral geniculate nuclei; LN, lateral nucleus; PN, parafascicular nucleus; VN, ventral thalamic nucleus; V, third ventricle. Arrow indicates the location of the bipolar stimulating electrode. *B*, antidromic invasion of VN cell after midline stimulation. *C*, synaptic activation from the same locus after increase in stimulus strength. The voltage and time calibration are indicated.

current pulse was superimposed on a d.c. depolarization, the cell fired repetitively. This is expected if the sum of the two depolarizations reaches the firing level of the cell. However, when the membrane potential was hyperpolarized from rest, another type of electroresponsiveness was observed. Indeed, the excitability of the cell was increased by membrane hyperpolarization. Thus, single burst responses having an abrupt onset and a smooth falling phase were obtained by current pulses as in Fig. 2*A* when the cell was previously hyperpolarized by 7 mV (Fig. 2*B*). Note the marked difference between the firing level for the tonic firing (arrow in Fig. 2*A*) and that of the burst response (arrow in Fig. 2*B*). The burst response had two distinct components: (1) a slowly rising all-or-none depolarization, the amplitude of which is directly related to the level of membrane hyperpolarization; and (2) one or more fast spikes activated from the slow depolarization.

Such a burst was also observed as a rebound response at the break of a hyperpolarizing current pulse (Fig. 2*D*). In these conditions the two components could be easily separated. In fact, with low-amplitude hyperpolarizing pulses only the slow component was observed, emphasizing the graded nature of the amplitude of this response. With larger hyperpolarizations the slow component reached the threshold for fast spike activation. Note that the hyperpolarization had a slow recovery time suggesting the presence of an 'early Na conductance'. The increased excitability that occurred with hyperpolarization was also demonstrated after synaptic activation. Figure 2*E* shows that smoothly graded excitatory synaptic potentials which reached the firing level were obtained in a thalamic neurone having a resting potential of −60 mV after midline thalamic stimulation of increasing amplitude. When the neurone was hyperpolarized by 15 mV (Fig. 2*F*), the same synaptic activation generated all-or-none slow depolarization, which in turn activated fast spikes. Note once again the different firing levels for the activation of the slow response and the fast action potentials.

These two modes of electroresponsiveness were found to have different ionic mechanisms as demonstrated by the addition of specific ionic channel blockers to the bathing solution. The fast action potentials, which seem to arise from the soma and axon areas (judging from the extracellular negativity they generate) can be blocked by tetrodotoxin (TTX) indicating that they are sodium-dependent[8]. As shown in Fig. 2*G*, a step

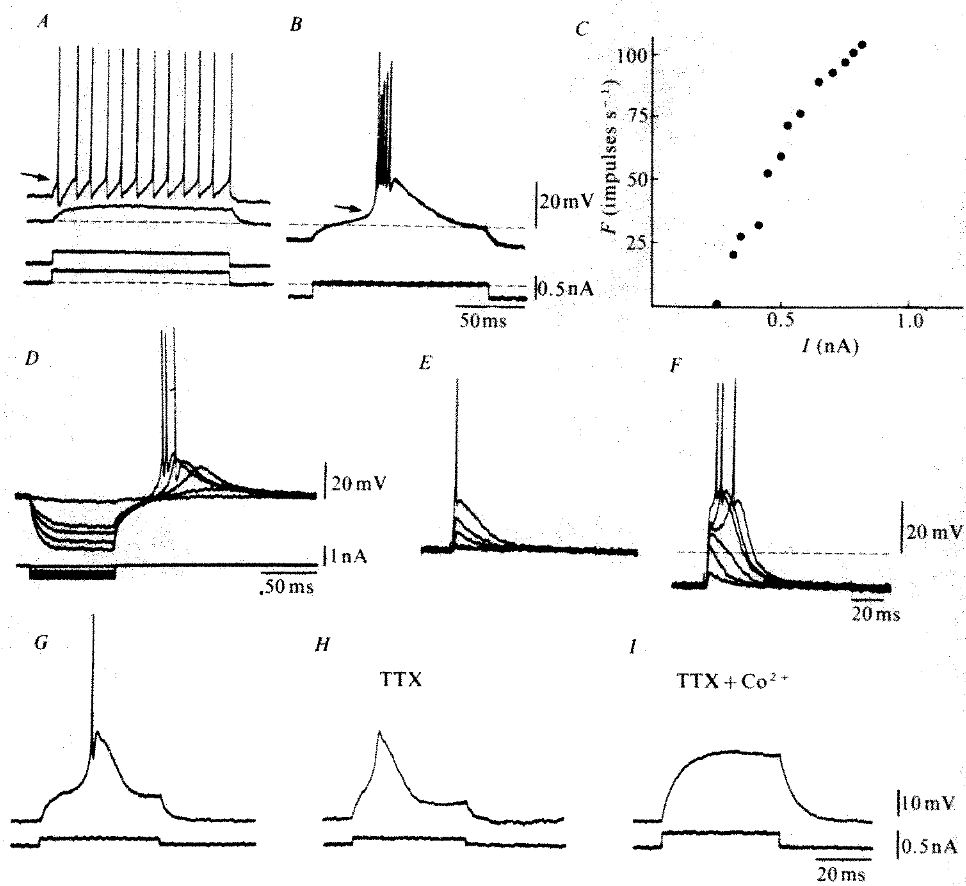


Fig. 2 Electrophysiological and pharmacological properties of thalamic neurones. *A*, a subthreshold depolarization of the cell at resting level (broken line) produces, after a d.c. depolarization, repetitive firing of the cell during the same current pulse. *B*, after d.c. hyperpolarization, similar current pulses as in *A* produce a single high frequency spike burst. *C*, plot of the instantaneous adapted frequency (f) of the repetitive firing of the cells as seen in *A* for different levels of current injection (I). *D*, rebound burst response after hyperpolarizing pulses of different amplitudes. *E*, in another cell, excitatory synaptic potentials obtained from rest show their graded character and the firing level of the neurone. *F*, synaptic potentials of similar amplitude as in *E* produce, following hyperpolarization, long-lasting all-or-none responses on which fast spikes are generated at the same firing level as in *E*. *G*, slow all-or-none response generating fast spike, from a cell slightly hyperpolarized from rest potential. *H*, after blockage of Na^+ conductance with TTX, the fast action potential is blocked but the slow response remains unchanged. *I*, addition of $CoCl_2$ to the bathing solution abolishes completely the slow response seen in *H*, even when the current pulse is increased in amplitude by 2.5 times.

depolarization from a holding level of −65 mV produced a large all-or-none depolarization and a secondary fast spike. Addition of TTX to the bath at 1 μg ml^{-1} completely blocked the fast action potential, leaving the underlying all-or-none slow depolarization unaltered (Fig. 2*H*). By contrast, addition of Ca^{2+}-conductance blockers such as $CoCl_2$ (ref. 9) or $CdCl_2$ (ref. 10) to the bathing solution at a concentration of 1 mM, or substitution of $MnCl_2$ (3.5 mM) for $CaCl_2$, abolished this all-or-none response (Fig. 2*I*). These results indicate that the slow spike is generated by a voltage-dependent change in Ca^{2+} conductance[11]. Furthermore, the nature of this response is similar to the low-threshold Ca^{2+}-dependent spikes described in the inferior olive[5,12].

A more detailed study of the switching between a Ca^{2+}-dominated and a Na^{+}-dominated electroresponsiveness revealed that this change occurs in almost an all-or-none manner. Figure 3 shows that when the excitability of a cell was challenged with a series of brief depolarizing pulses (of sufficient amplitude to fire the cell), and simultaneously the membrane potential was gradually decreased by a slow ramp depolarizing current, the response to the pulse changed from the burst mode to tonic firing. In this case the bursts had a duration of 30–130 ms, and generated a set of fast action potentials at a frequency of 150–320 impulses s^{-1}, depending on the level of membrane hyperpolarization. As the membrane was slowly depolarized, the test pulses produced, rather than the burst, a continuous repetitive spiking for the duration of the pulse (compare *b* and *c* in Fig. 3*B*). The frequency of this repetitive firing was related to the total level of depolarization as indicated in the *f–I* plot in Fig. 2*C*.

The mechanism for the switching between these two excitability states was clarified by further studying the properties of the Ca^{2+}-dependent electroresponsiveness. As shown in Fig. 2*D*, the amplitude and the rate of rise of this Ca^{2+}-dependent response varies with membrane potential in a very steep manner. In addition, as in the case of the inferior olivary cells, these responses completely inactivate at membrane potential levels more positive than −60 mV. This voltage-dependent inactivation explains both the presence of burst responses when the cell is hyperpolarized and the rebound activation of these cells. Furthermore, as observed in inferior olivary neurones[5], these rebound calcium spikes show a long refraction, indicating that the kinetics of recovery from the inactivated state are slow. This slow recovery also explains why such responses are unitary as their activation is followed by a rather prolonged refractory period of 80–150 ms.

In conclusion, these findings indicate that thalamic neurones have special voltage-sensitive ionic conductance properties which allow them to change from a phasic bursting response (followed by neuronal silence) to graded repetitive firing, and this switching is modulated by membrane potential. When the cells have a resting level more negative than −60 mV, a low threshold calcium response is obtained. The other integrative state is characterized by tonic firing brought about by depolarization of the cell to levels more positive than −60 mV. At this membrane potential level, the calcium conductance which generates the slow response is inactivated and the cell fires repetitively. Although the precise distribution of these conductances over the soma-dendritic area of these neurones has not been determined, by analogy with inferior olivary cells[5,12], we suggest that the Ca^{2+}-deinactivated response probably resides at the somatic level, as does the Na^{+}-dependent action potential.

These electrical properties further explain many of the electroresponsive properties observed in thalamic recordings *in vivo* where such switching of tonic to phasic firing has in fact been described[13,14]. Moreover, in conjunction with the recurrent inhibition observed *in vivo*, this newly described Ca^{2+} conductance represents the functional basis of the so-called 'post-anodal exaltation' underlying the recruiting responses[12] at cortical level and most probably the α rhythm[15].

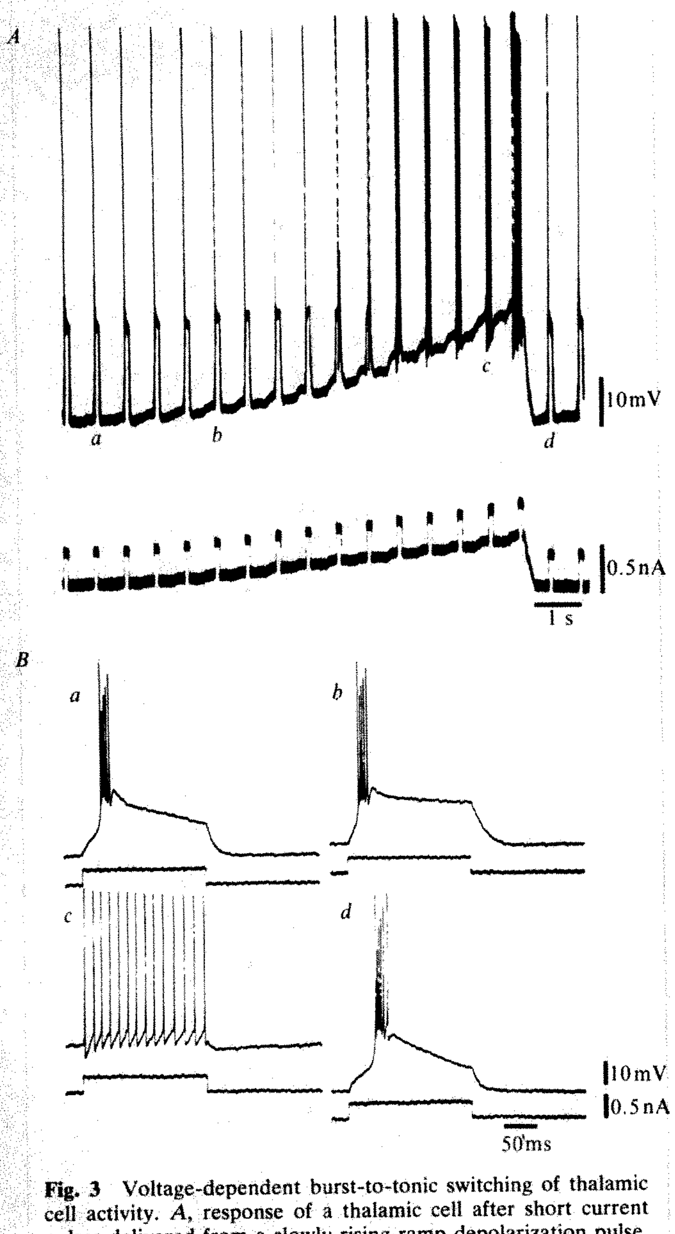


Fig. 3 Voltage-dependent burst-to-tonic switching of thalamic cell activity. *A*, response of a thalamic cell after short current pulses delivered from a slowly rising ramp depolarization pulse. The cell switched abruptly from a burst response to tonic firing as the d.c. potential decreased by ~10 mV from the initial value. *B*, records obtained at a higher sweep speed at the times indicated by *a* to *d* in *A*. Note the transition from burst response (*a* and *b*) to tonic response (*c*), followed by the abrupt return to a burst response (*d*).

This research was supported by USPHS grant NS13742 from the National Institute of Neurological and Communicative Disorders and Stroke. H.J. was supported in part by grants from the University of Copenhagen, the Danish MRC and the Weimann Foundation, and by an Albert Cass Travelling Fellowship.

Received 21 January; accepted 29 March 1982.

1. Shepherd, G. D. *The Synaptic Organization of the Brain*, 2nd edn (Oxford University Press, New York & Oxford, 1979).
2. Llinas, R. & Sugimori, M. *J. Physiol., Lond.* **305**, 171–195 (1980).
3. Granit, R. *Mechanisms Regulating the Discharge of Motoneurons* (Charles C. Thomas, Illinois, 1972).
4. Fulton, B. P., Miledi, R. & Takahashi, T. *Proc. R. Soc.* **B208**, 115–120 (1980).
5. Llinas, R. & Yarom, Y. *J. Physiol., Lond.* **315**, 549–567 (1981).
6. Schwartzkroin, P. A. *Brain Res.* **85**, 423–436 (1975).
7. Connor, J. A. & Stevens, C. F. *J. Physiol., Lond.* **213**, 21–30 (1971).
8. Narahashi, T., Moore, J. W. & Scott, W. R. *J. gen. Physiol.* **47**, 965–974 (1964).
9. Baker, P. F., Hodgkin, A. L. & Ridgway, E. G. *J. Physiol., Lond.* **218**, 709–755 (1971).
10. Kostyuk, P. G. & Krishtal, O. A. *J. Physiol., Lond.* **270**, 545–568 (1977).
11. Hagiwara, S. *Adv. Biophys.* **4**, 71–102 (1973).
12. Llinas, R. & Yarom, Y. *J. Physiol., Lond.* **315**, 569–581 (1981).
13. Purpura, D. P. & Cohen, B. *J. Neurophysiol.* **25**, 621–635 (1962).
14. Maendly, R. *et al. J. Neurophysiol.* **46**, 901–917 (1981).
15. Andersen, P., Eccles, J. C. & Sears, T. A. *J. Physiol., Lond.* **174**, 370–399 (1964).

Evidence for the presence of S-100 protein in the glial component of the human enteric nervous system

Gian-Luca Ferri*, Lesley Probert*, Domenico Cocchia†, Fabrizio Michetti†, Paul J. Marangos‡ & Julia M. Polak*

* Department of Histochemistry, Royal Postgraduate Medical School, Hammersmith Hospital, London W12 0HS, UK
† Department of Anatomy, Università Cattolica, 00168 Roma, Italy
‡ Clinical Psychobiology Branch, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland 20205, USA

Recently, the complex organization of the enteric nervous system and its independent role in the integrative control of digestive functions have been fully recognized. However, little is known about the non-neuronal elements present in close morphological and, presumably, functional relationship with the enteric neurones. Although 'enteric glial cells' in gut ganglia and Schwann cells accompanying nerve fibres throughout the gastrointestinal wall can easily be recognized by electron microscopy on the basis of their ultrastructural features[1–3], the lack of a suitable common marker for the visualization of these non-neuronal components of the enteric nervous system has hampered investigation. S-100 is a Ca^{2+}-binding protein[4] originally isolated from the brain[5], where it is localized mainly in glia[6,7]. We demonstrate here, using immunohistochemistry and electron immunocytochemistry, that both enteric glial cells and Schwann cells of the human gut contain densely immunoreactive S-100. This protein can therefore be regarded as a common marker for the glial components of the enteric nervous system.

The distribution of the S-100 protein in the enteric nervous system was studied by immunocytochemistry in tissue samples obtained from uninvolved areas of segments of human gut resected for carcinoma (stomach, ileum and colon; $n = 15$). Throughout the entire length of the gastrointestinal tract, S-100-immunoreactivity was detected in the ganglia of both the submucous and the myenteric plexus, in slender cells having numerous projections partially surrounding enteric neurones (Fig. 1*a*), presumably glial cells. In addition, numerous elongated, dendritic cellular elements were immunostained in all non-epithelial layers of the wall (Fig. 2); these were considered to be Schwann cells.

To confirm that in enteric ganglia S-100 is selectively localized in glial cells, we compared the distribution of this antigen with that of neurone-specific enolase (NSE), an isomer of the glycolytic enzyme enolase that is specifically localized in neurones of the central nervous system (CNS)[7,8] and has also been demonstrated in gut neurones[9]. Figure 1 shows that S-100 was localized only in glial cells (*a*), while neurones, identified by NSE immunostaining (*b*), were completely negative.

Ultrastructurally identifiable Schwann cells accompany the bundles of nerve fibres running through all layers of the gut wall, forming an incomplete sheath around them[2]. To confirm that the S-100-immunoreactive elements found outside the enteric ganglia were Schwann cells, we studied their three-dimensional structure and distribution in stretch preparations of either submucosa or lamina propria separated by microdissection, then immunostained and examined whole[10]. In both of these layers, a network of S-100-immunoreactive elements was demonstrated in the internodal strands running between enteric ganglia (Fig. 3*a*), around blood vessels, and in a dense mesh of interconnecting cells both at the level of the muscularis mucosae (Fig. 3*b*) and in the lamina propria up to close contact with the basement membrane of the epithelium. The distribution of S-100-immunoreactivity seems therefore to parallel the general distribution of the enteric innervation.

Glial cells in enteric ganglia have distinctive ultrastructural features resembling those of astrocytes in the CNS[3]. Typical

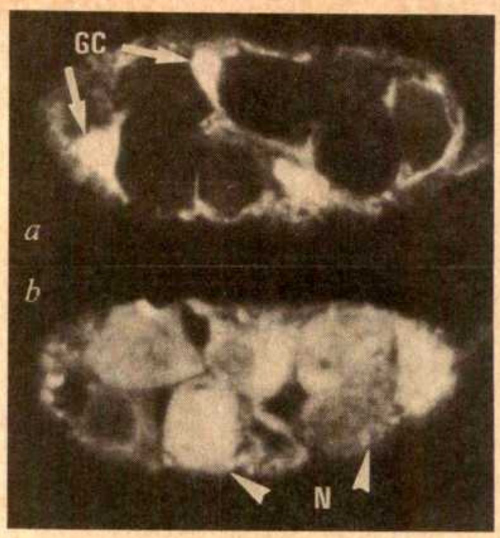


Fig. 1 Immunofluorescent detection of S-100 (*a*) and NSE (*b*) in two serial (4 μm) sections of a submucous ganglion in the human colon. Tissues were fixed in *p*-benzoquinone solution[21] (0.4% in phosphate-buffered saline, PBS) for 2 h, washed overnight in PBS containing 7% sucrose, frozen in melting Arcton and cut with a cryostat. Sections were stained by indirect immunofluorescence using anti-S-100 antiserum diluted 1:500, and anti-NSE antiserum diluted 1:800 in PBS (overnight incubation), followed by fluorescein isothiocyanate-conjugated goat anti-rabbit IgG diluted 1:200 (1 h). Specific rabbit antisera raised against ox S-100[14] and rat NSE[8], fully cross-reacting with the human S-100 and NSE, were used. Controls included the use of antibodies preabsorbed with 20 $\mu g\ ml^{-1}$ of ox S-100 and 5 $nmol\ ml^{-1}$ of rat or human NSE. S-100 immunoreactivity (*a*) was confined to the slender and elongated glial cells (GC) partially surrounding neurones (N), which were completely negative for this antigen and were revealed (*b*) by NSE immunostaining. ×700.

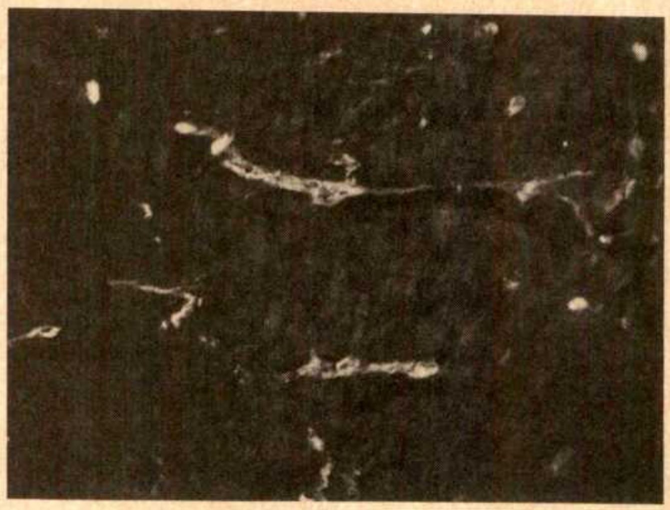


Fig. 2 S-100 immunofluorescence (fixation and immunostaining as described in Fig. 1 legend) in the muscle layer of the human colon. Numerous slender cells having several long, branching processes can be observed. Similar elements, with the morphological appearance of Schwann cells, were demonstrated in all non-epithelial layers of the gut wall (10 μm section, ×120).

0028-0836/82/220409—02$01.00

glial cells identified by electron microscopy in enteric ganglia contain S-100-immunoreactivity, while enteric neurones are negative (Fig. 4*a*). In addition, in all gut layers, positive immunostaining for S-100 was demonstrated in Schwann cells partially surrounding bundles of nerve fibres (Fig. 4*b*). In the present experiments, only the cytoplasmic localization was consistently demonstrated in enteric glia, even though a nuclear localization of this protein has been documented in the CNS[11-13].

Although S-100 has recently been demonstrated in various locations outside the nervous system, including stellate cells in the adenohypophysis[14,15], satellite cells in the adrenal medulla[16], melanocytes and Langerhans cells in the skin[17] and chondrocytes[18], there is general agreement on its prevalent localization in the CNS in glia[6,7]. The presence of this brain protein in the enteric nervous system agrees with the numerous homologies reported between the gut 'mini brain' and the CNS[2].

Another brain protein, the glial fibrillary acidic protein, has recently been demonstrated in enteric glial cells[19]. However, the absence of this protein from Schwann cells[19] precludes its use as a common marker of the non-neuronal components of the enteric nervous system.

In conclusion, our results indicate that the S-100 protein can be reliably used to demonstrate the glial elements of the enteric nervous system in man. This will have immediate relevance to the study of disease conditions affecting the gastrointestinal innervation, especially those in which an abnormal glial proliferation is suggested[20].

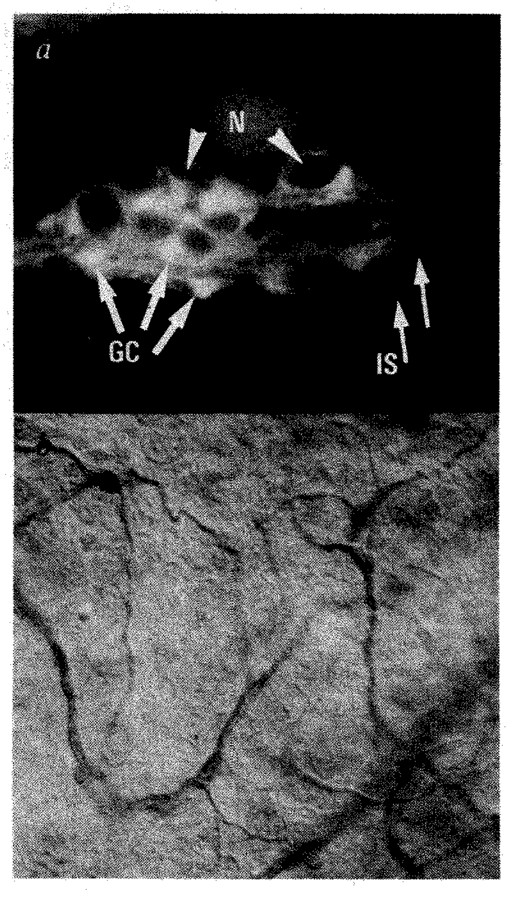


Fig. 3 S-100 immunoreactivity in whole-mount preparations of submucosa[10]. This layer was separated from samples of ileum and colon by microdissection, fixed in *p*-benzoquinone (see Fig. 1 legend), dehydrated through graded alcohols, cleared in xylene, rehydrated and immunostained *in toto* by immunofluorescence or by the peroxidase-antiperoxidase technique[22]. Prolonged incubations (16–36 h) and washings (6–8 h) were used. S-100 immunostaining revealed the general distribution of glial elements in the whole enteric innervation; *a*, in glial cells (GC) partially circumscribing unstained neurones (N) in the ganglia, and in Schwann cells in internodal strands (IS); and *b*, in a rich network of Schwann cells at the level of the muscularis mucosae. (*a*, ×375; *b*, ×280.)

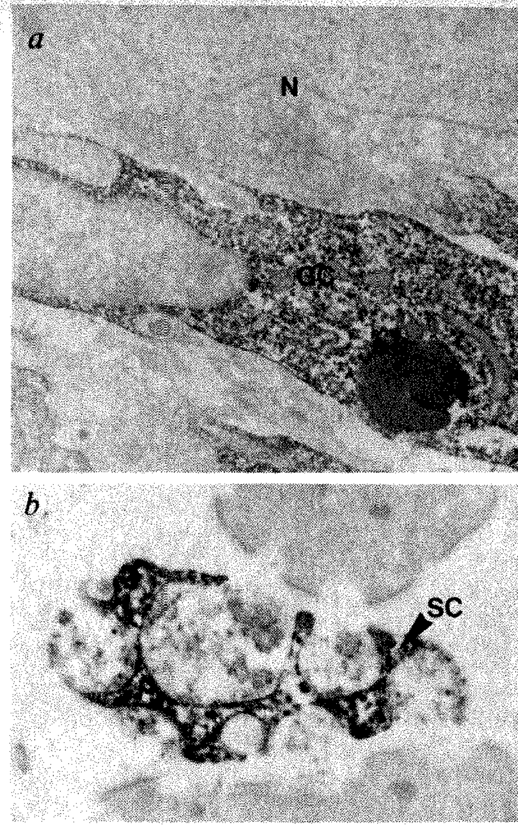


Fig. 4 Electron immunocytochemistry for S-100. Tissues were fixed in 4% paraformaldehyde in 0.1 M phosphate buffer for 3 h, washed overnight in PBS and cut at 100 μm with a vibratome. Sections were stained by the peroxidase–antiperoxidase technique[14,23] (same antibody dilution and controls as described for Fig. 1) using diaminobenzidine acid as coupler, then embedded in Araldite and cut at 60 nm for electron microscopy. In enteric ganglia (*a*), immunoreactivity is demonstrated in the cytoplasm of glial cells (GC), which partially circumscribe neurones that are completely unstained (N). In a nerve bundle in the muscle layer (*b*), projections of Schwann cells (SC) are in close contact with nerve fibres, which are unstained (ileum: *a*, ×10,000; *b*, ×20,000).

G.-L.F. is a research fellow from the Department of Medicine 1, University of Bologna, Italy, receiving financial support from the Accademia Nazionale dei Lincei, L.P. is a MRC student.

Received 5 February; accepted 7 April 1982.

1. Cook, R. D. & Burnstock, G. *J. Neurocytol.* **5**, 195–206 (1976).
2. Gabella, G. *Int. Rev. Cytol.* **59**, 129–193 (1979).
3. Gabella, G. *Neuroscience* **6**, 425–436 (1981).
4. Klee, C. B., Crouch, T. H. & Richman, P. G. *A Rev. Biochem.* **49**, 489–511 (1980).
5. Moore, B. W. *Biochem. biophys. Res. Commun.* **6**, 739–744 (1965).
6. Bock, E. *J. Neurochem.* **30**, 7–14 (1978).
7. Zomzely-Neurath, C. E. & Walker, W. A. in *Proteins of the Nervous System* (eds Bradshaw, R. A. & Schneider, D. M.) 1–57 (Raven, New York, 1980).
8. Marangos, P. J., Zomzely-Neurath, C., Luk, D. C. M. & York, C. *J. biol. Chem.* **250**, 1884–1891 (1975).
9. Tapia, *et al. Gut* **21**, 920 (1980).
10. Ferri, G-L. *et al. Gut* **22**, 898 (1981).
11. Michetti, F., Miani, N., De Renzis, G., Caniglia, A. & Correr, S. *J. Neurochem.* **22**, 239–244 (1974).
12. Ludwin, S. K., Kosek, J. C. & Eng, L. F. *J. comp. Neurol.* **165**, 197–208 (1976).
13. Cocchia, D. *Cell Tissue Res.* **214**, 529–540 (1981).
14. Cocchia, D. & Miani, N. *J. Neurocytol.* **9**, 771–782 (1980).
15. Nakajima, T., Yamaguchi, H. & Takahashi, K. *Brain Res.* **191**, 523–531 (1980).
16. Cocchia, D. & Michetti, F. *Cell Tissue Res.* **215**, 103–112 (1981).
17. Cocchia, D., Michetti, F. & Donato R. *Nature* **294**, 85–87 (1981).
18. Stefansson, K., Wollmann, R. L., Moore, B. W. & Arnason, B. G. W. *Nature* **295**, 63–64 (1982).
19. Jessen, K. R. & Mirsky, R. *Nature* **286**, 736–737 (1980).
20. Smith, B. *The Neuropathology of the Alimentary Tract* (Arnold, London, 1972).
21. Bishop, A. E., Polak, J. M., Bloom, S. R. & Pearse, A. G. E. *J. Endocr.* **77**, 25–26 (1978).
22. Sternberger, L. *Immunocytochemistry*, 2nd edn (Wiley, New York, 1979).
23. Pickel, V. M., Joh, H. T. & Reis, D. J. *Proc. natn. Acad. Sci. U.S.A.* **72**, 659–663 (1975).

Hapten-specific T suppressor factor recognizes both hapten and I–J region products on haptenized spleen cells

M. Zembala*, G. L. Asherson & V. Colizzi*

Division of Immunological Medicine, Clinical Research Centre, Watford Road, Harrow HA1 3UJ, UK

The classical example of genetic restriction is associative recognition, the finding that cytotoxic T-cell killing requires matching between part of the major histocompatibility complex (MHC) of the stimulating cell and the target cell[1,2]. Its analysis would be facilitated by the study of a system in which antigen-specific T suppressor factor (TSF) was genetically restricted in its interaction with haptenized cells. However, only one such system is known[3,4], although there are several systems in which the TSF must be syngeneic with other cells[5–8]. We now describe a genetically restricted TSF which arms an acceptor T cell. This acceptor cell shows some resemblance to the auxiliary T suppressor (Ts) cell, the acceptor T cell and the Ts_3 described elsewhere[5,6,9]. In the presence of antigen (haptenized spleen cells) the armed acceptor cell releases a nonspecific inhibitor of the passive transfer of contact sensitivity (see. Fig. 1). The important finding is that the antigen will only trigger the release of the nonspecific inhibitor when there is matching in the *I–J* region between the TSF and the haptenized cell used as a source of the triggering antigen.

In practice, the T acceptor (T_{acc}) cell is produced by immunization with picryl chloride (pic) or oxazolone. The contact sensitizer used is unimportant because it does not influence the specificity of the system (Table 1). Macrophages are removed by nylon wool filtration as they may act as acceptor cells[10]. The T_{acc} cell population is armed by incubation in anti-pic or anti-oxazolone TSF[11,12]. The cells are then exposed to antigen (picrylated spleen cells) and the nonspecific inhibitor released into the supernatant is tested by its ability to block the transfer of contact sensitivity by immune cells incubated in it. This system is formally analogous to the reaginic system in which the roles of specific IgE, mast cell, soluble antigen and histamine are taken by specific TSF, the acceptor T-cell, haptenized spleen cells and the nonspecific inhibitor (see Fig. 1).

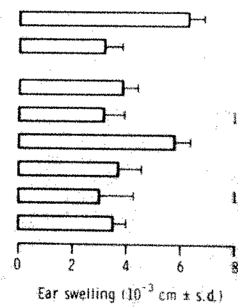
The first experiment investigated the main characteristics of the system. T acceptor cells were produced by immunization with picryl chloride or the unrelated contact sensitizer oxazolone, and were armed with anti-picryl TSF. After washing, picrylated spleen cells were added as a source of antigen and the supernatant was collected after 2 h. Table 1 (lines 3–5) shows that nonspecific inhibitor of the passive transfer of contact sensitivity was produced when the T_{acc} cells were generated by painting with picryl chloride or oxazolone, while normal spleen cells were ineffective, that is, immunization but not specific immunization was required for the generation of T_{acc} cells. Unseparated or nylon wool-purified T spleen cells could be used as a source of antigen (lines 3, 7) and irradiation at 2,000 rad did not affect the ability of picrylated spleen cells to trigger the release of nonspecific inhibitor.

We next investigated the requirement for genetic matching between the anti-picryl TSF and the picrylated cells used as a source of antigen. Immune T cells from CBA or BALB/c mice were used as an acceptor cell population and armed with anti-picryl TSF from CBA or BALB mice. Picrylated CBA or BALB spleen cells were then added as a source of antigen and the supernatant was collected at 2 h. Table 2 (line 3) shows that nonspecific inhibitor of the passive transfer of contact sensitivity was detected when the TSF, the acceptor T cell, the antigen (picrylated spleen cell) and the final test cell were all syngeneic. Lines 7, 9 and 11 of Table 2 show that syngeneity between the TSF and the antigen (picrylated spleen cell) was critical for triggering the release of the nonspecific inhibitor from the acceptor cell.

* Permanent addresses: Institute of Paediatrics, Copernicus School of Medicine, Wielicka 265, 30-663 Kracow, Poland (M.Z.); Institute of Microbiology, University of Pisa, 56100 Pisa, Italy (V.C.).

Table 1 Acceptor cells require immunization but not specific immunization for their production

	TSF	Antigen	T_{acc}	Test cell	
1	Positive control			Ox CBA	
2	Negative control			-	
3	Pic CBA	Pic CBA	Pic CBA	Ox CBA	77% ***
4	″	″	Ox CBA	″	100% ***
5	″	″	- CBA	″	20%
6	″	″ †	Pic CBA	″	81% **
7	″	T_{nyl} pic CBA	″	″	100% ***
8	″	″ †	″	″	98% ***



Anti-picryl TSF was produced by injecting picrylsulphonic acid (PSA) intravenously, painting with picryl chloride on day 6 and culturing lymphoid cells taken on day 7 for 48 h (ref. 12). Nylon wool-purified T spleen cells (T_{nyl}, 1.5×10^8) taken from mice 4 days after painting with picryl chloride (Pic) or oxazolone (Ox) or from normal mice were used as acceptor cells. They were armed with anti-picryl TSF (4 ml) at 37 °C for 1 h, spun down, washed once and antigen added. The standard antigen was picrylated normal spleen cells (3×10^7), prepared by lysis of red cells with Boyle's solution and picrylation with 10 mM PSA for 10 min (ref. 22). The armed acceptor cells and the picrylated spleen cells were centrifuged together and resuspended in 5 ml Eagle's minimal essential medium/Dulbecco's phosphate-buffered saline (MEM/PBS) with 2.5% inactivated foetal calf serum and incubated for 2 h with shaking every 40 min. Finally the nonspecific inhibitor in the supernatant was tested by incubating passive transfer cells (2.5×10^8, 4 day, oxazolone immune lymph node and spleen cells) with nonspecific supernatant (5 ml) for 1 h, washing once and injecting 4.5×10^7 into each of four to five CBA mice[12]. The mice were immediately challenged with 1% oxazolone on both sides of both ears and contact sensitivity was assessed by the ear swelling at 24 h. The horizontal bars show the increase in ear thickness at 24 h in units of 10^{-3} cm ± s.d. The figures show the percent depression of contact sensitivity taking the positive control group incubated in medium as 0% and the negative control group which did not receive any cells as 100%. †The picrylated spleen cells used as antigen were irradiated at 2,000 rad from a cobalt source. The significance of the difference from the positive control by Student's *t*-test was $^*P < 0.05$, $^{**}P < 0.002$ and $^{***}P < 0.001$.

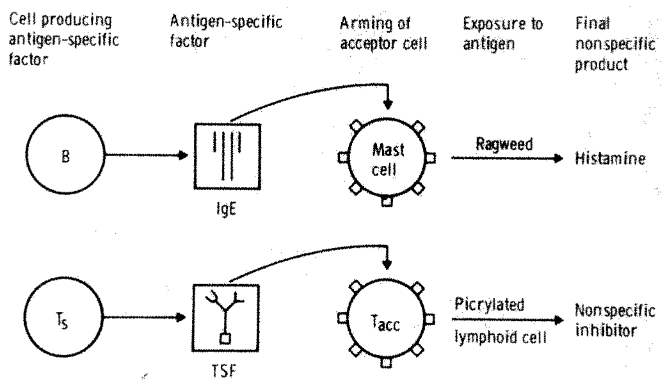
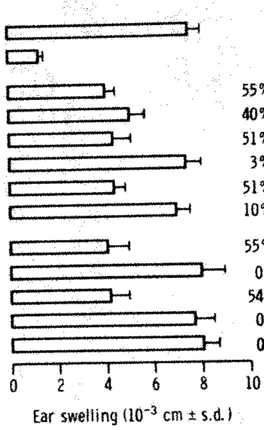


Fig. 1 Analogy between the IgE mast cell system and the T suppressor circuit involving Ts-eff, T suppressor factor and the T acceptor cell. In allergy to ragweed, the B cell makes specific IgE antibody. This is cytophilic and behaves like a mobile receptor which arms the mast cell. On exposure to ragweed the mast cell releases the final nonspecific products such as histamine. Similarly, in the T suppressor system to the picryl group, the T suppressor cell produces an anti-picryl T suppressor factor. This is cytophilic and arms the T acceptor cell. On exposure to antigen (picrylated lymphoid cells with the same *I–J* region as the cells producing the TSF) the T acceptor cell releases the final nonspecific products such as the inhibitor of the transfer of contact sensitivity. The diagram of the T suppressor factor illustrates its binding sites for the attachment to the T acceptor cell (□), the recognition of the picryl hapten (<) and the recognition of the product of the *I–J* locus (⊂).

0028–0836/82/220411—03$01.00

Table 2 Acceptor cells only release nonspecific inhibitor when there is genetic matching between the anti-picryl T suppressor factor and the triggering antigen (picrylated spleen cells)

	TSF	Antigen	T_{acc}	Test cell	
1	Positive control			Ox CBA	
2	Negative control			-	
3	CBA	CBA	CBA	Ox CBA	55%***
4	"	"	" †	"	40%***
5	"	"	" ††	"	51%***
6	"	BALB	"	"	3%
7	BALB	"	"	"	51%***
8	"	CBA	"	"	10%
9	BALB	BALB	BALB	Ox CBA	55%***
10	"	CBA	"	"	0%
11	CBA	CBA	"	"	54%***
12	"	BALB	"	"	0%
13	-	"	"	"	0%



Pooled lymph node and spleen T cells from mice painted with picryl chloride were used as acceptor cells and armed with anti-picryl TSF. They were exposed to antigen (picrylated spleen cells) and the nonspecific inhibitor was tested on the passive transfer of contact sensitivity to oxazolone in CBA mice. The experiment studied the effect on the generation of nonspecific inhibitor of the genotype (CBA or BALB/c) of the T_{acc} cell source, the TSF and the spleen cells used as a source of picrylated antigen.

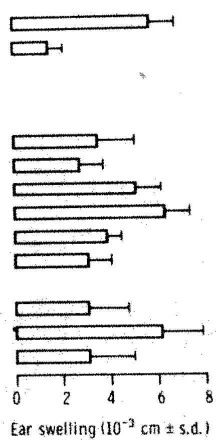
†, Unhaptenized CBA spleen cells (line 4) and ††, unhaptenized BALB cells (line 5) were added (3×10^7) to show that BALB cells did not block the liberation of nonspecific inhibitor through an allogeneic effect.

Congeneic B10.A mice were then used to confirm that the MHC was responsible for the genetic restriction. Anti-picryl TSF was prepared in B10.A and B10.A(3R) mice and used to arm CBA acceptor cells. Various picrylated cells were then used to trigger the release of nonspecific inhibitor. Table 3 shows that syngeneity at *I–J* was both necessary and sufficient.

The question was then asked whether the same molecule in the TSF preparation both had a combining site for antigen and was responsible for the genetic restriction. In a preliminary experiment, TSF purified by absorption onto haptenized microcrystalline cellulose and elution with the corresponding haptenized ε-aminocaproic acid showed genetic restriction. To confirm this result, anti-picryl and anti-oxazolone TSF were purified by binding to the appropriate haptenized microcrystalline cellulose. Dummy elution was then undertaken with the 'wrong' hapten followed by elution with the right hapten. Table 4 shows that the dummy eluates were inactive while the active TSF eluted by the appropriate hapten showed genetic restriction. We concluded that the hapten recognition site and the site which recognized the product(s) of the *I-J* locus of the MHC occurred on the same molecule. The fact that the effect of arming T_{acc} cells with TSF persists after washing suggests that TSF has a binding site for such cells. Stronger evidence is provided by the observation that T_{acc} cells armed with TSF can be purified by adherence (panning) on Petri dishes coated with haptenized albumin[13]. Hence, the TSF has three binding sites—one for attachment to the acceptor T cell, one for recognition of antigen and one for recognition of the products of the *I-J* locus (see model of TSF in Fig. 1). Feldmann[14] has suggested a similar schema. In fact, in the present system, questions about the nature of the combining site(s) involved in associative recognition become structural questions about antigen-specific TSF.

Table 3 Matching in the *I–J* region between anti-picryl T suppressor factor and the triggering antigen (picrylated spleen cells) is required for the release of nonspecific inhibitor

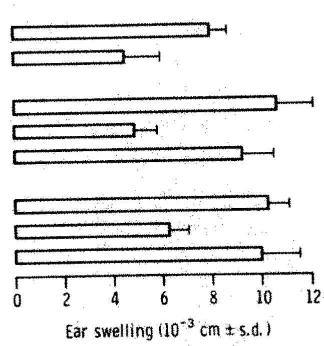
TSF	ag	K	Ia: A	Ia: B	Ia: J	Ia: E	C	D	
	Positive control								
	Negative control								
B10.A	B10.A	+	+	+	+	+	+	+	53%*
	2R	+	+	+	+	+			73%***
	3R					+	+	+	14%
	4R	+	+						0%
	5R				+	+			46%***
	CBA	+	+	+	+	+			64%*
3R	3R	+	+	+	+	+	+	+	64%*
	5R	+	+	+		+	+	+	0%
	4R			+	+				65%*



Anti-picryl TSF was prepared in B10.A and B10.A (3R) mice and used to arm CBA acceptor cells produced by immunization with picryl chloride. Picrylated spleen cells from congeneic mice were used to trigger the release of nonspecific factor and the supernatant was tested for its ability to inhibit the passive transfer of contact sensitivity to oxazolone using CBA cells. The plus signs refer to syngeneity between the TSF and the antigen used for triggering the release of nonspecific inhibitor.

Table 4 Purified anti-picryl and anti-oxazolone T suppressor factors are genetically restricted in their action

TSF	Elution	Antigen	
Positive control			
Negative control			
Pic CBA	Ox-EACA	CBA	0%
	Pic-EACA	CBA	89%***
	Pic-EACA	BALB	0%
Ox CBA	Pic-EACA	CBA	0%
	Ox-EACA	CBA	49%***
	Ox-EACA	BALB	0%



Anti-picryl and anti-oxazolone TSF[12] (15 ml) were absorbed for 2 h with 0.5 ml picrylated or oxazolonated diaminohexane microcrystalline cellulose prepared by haptenizing beads coupled to diaminohexane using sodium periodate and sodium borohydride. They were then eluted with the wrong haptenized ε-aminocaproic acid (3 mg ml^{-1}) for 45 min and then with the appropriate hapten. The eluates were made 0.25% with inactivated newborn calf serum and dialysed against PBS and MEM/PBS. Acceptor cells produced by immunizing CBA mice with picryl chloride were armed with the supernatant. Antigen (CBA or BALB picrylated spleen cells) was then added and the supernatant tested for its ability to inhibit the passive transfer of contact sensitivity to oxazolone using CBA cells.

It is puzzling that the *I-J* locus and the several *I-J* determinants[15] are involved at multiple levels of T suppressor systems. In particular, the TSF is probably made by an I-J^+ cell and TSF, including anti-picryl TSF but not anti-dinitrophenyl TSF[3], carries *I-J* determinants[16–18]. The T_{acc} cell studied here and those reported elsewhere[5,6,9] are I-J^+ and the genetic restriction between the TSF and the triggering antigen in the present system is *I-J*-restricted. An *I*-region restriction between antigen-specific TSF which induces an anti-idiotype Ts_2 and the Ts_3 is also described[16]. Finally in neonatal tolerance *I-J* differences favour unresponsiveness[19] and the presence of allogeneic *I-J* augments the activity of certain immune T suppressor cells[20].

Czitrom, Sunshine and Mitchison[21] proposed that antigen presented in the context of *I-J*, that is, on an *I-J*-positive macrophage or *T* cell[15], selectively stimulates T suppressor cells. This implies that the T_s and the TSF produced by the T_s carry receptors both for antigen and for the *I-J* on the antigen-presenting cell. This hypothesis would explain the requirement for syngeneity between the mouse producing the TSF, in which

antigen was initially presented, and the antigen used to trigger the release of nonspecific inhibitor. An interesting elaboration is that the ability of antigen-specific TSF to give rise to anti-idiotype T suppressor cells[6] might be due to the *I–J* antigen of the TSF (or of the macrophage or T acceptor cell to which it attaches) acting as a signal for the induction of suppressor cells.

M.Z. and V.C. thank the Wellcome Trust and the Lega Nazionale contro i Tumori for financial support, Malcolm Brenner, Miroslav Malkovsky and Liz Simpson for valuable discussion and Valerie Stein and Madelaine Watkins for technical assistance.

Received 29 January; accepted 26 March 1982.

1. Zinkernagel, R. M. & Doherty, P. C. *Adv. Immun.* **27,** 52–177 (1979).
2. Shearer, G. M., Remn, T. G. & Garbarino, C. A. *J. exp Med.* **141,** 1348–1364 (1975).
3. Moorhead, J. W. *J. exp. Med.* **150,** 1432–1447, (1979).
4. Claman, N. H., Miller, S. D., Conlon, P. J. & Moorhead, J. W. *Adv. Immun.* **30,** 121–157, (1980).
5. Taniguchi, M. & Tokuhisa, T. *J. exp. Med.* **151,** 517–527 (1980).
6. Sunday, M. E., Benacerraf, B. & Dorf, M. E. *J. exp. Med.* **153,** 811–822 (1981).
7. Germain, R. N. & Benacerraf, B. *Springer Semin. Immunopath.* **3,** 93–128 (1980).
8. Germain, R. N. & Benacerraf, B. *Scand. J. Immun.* **13,** 1–10 (1981).
9. Sy, M. S., Miller, S. D., Moorhead, J. W. & Claman, H. N. *J. exp. Med.* **149,** 1197–1207 (1979).
10. Ptak, W., Zembala, M., Asherson, G. L. & Marcinkiewicz, J. *Int. Archs. Allergy appl. Immun.* **65,** 121–128 (1981).
11. Asherson, G. L., Zembala, M., Thomas, W. R. & Perera, M. A. C. C. *Immun. Rev.* **50,** 3–45 (1980).
12. Asherson, G. L., & Zembala, M. *Immunology* **42,** 1005–1013 (1980).
13. Asherson, G. L. & Zembala, M. *Ann. N. Y. Acad. Sci.* (in the press).
14. Feldmann, M. & Kontiaienen, S. *Molec. cell. Biochem.* **30,** 177–193 (1980).
15. Murphy, D. B., Yamauchi, K., Habu, S., Eardley, D. D. & Gershon, R. K. *Immunogenetics* **13,** 205–213 (1981).
16. Okuda, K., Minami M., Sher, D. H. & Dorf, M. E. *J. exp. Med.* **154,** 468–479 (1981).
17. Taniguchi, M., Takei, I. & Tada, T. *Nature* **283,** 227–228 (1980).
18. Noonan, F. P. & Halliday, W. J. *Cell. Immun.* **50,** 41–47 (1980).
19. Streilein, J. W. & Klein, J. *Proc. R. Soc.* B **207,** 461–474 (1980).
20. Bromberg, J. S., Benacerraf, B. & Greene, J. *J. exp. Med.* **153,** 437–449 (1981).
21. Czitrom, A. A., Sunshine, G. H. & Mitchison, N. A. *Immunogenetics* **11,** 97–102 (1980).
22. Mishell, B. B. & Shiigi, S. M. *Selected Methods in Cellular Immunology,* 98 (Freeman, San Francisco, 1980).

Epstein–Barr virus-specific cytotoxic T-cell clones restricted through a single HLA antigen

L. E. Wallace, A. B. Rickinson, M. Rowe & M. A. Epstein

Department of Pathology, University of Bristol Medical School, University Walk, Bristol BS8 1TD, UK

HLA-restriction of antiviral cytotoxic T-cell responses in man, precisely analogous to the H-2-restriction first documented in murine virus infections[1–3], has now been unequivocally demonstrated with influenza virus-specific [4–7] and with Epstein–Barr (EB) virus-specific[8–12] cytotoxic T cells reactivated *in vitro*. If individual effector cells of this kind recognize viral antigen in association with one particular HLA determinant, then it follows that cytotoxic T-cell preparations from HLA-heterozygous donors must be polyclonal with discrete component populations each functionally linked to one of the relevant HLA-A or -B antigens. Clonal expansion of individual cells from such preparations in medium containing T-cell growth factor (TCGF-medium) not only offers a means of testing this hypothesis but should also provide monospecific populations ideal for the analysis of human cytotoxic T-cell function. We describe here the establishment *in vitro* of EB virus-specific human cytotoxic T-cell clones each recognizing the virus-induced lymphocyte-detected membrane antigen (LYDMA) in association with one particular HLA-A or -B determinant.

EB virus-specific cytotoxic T cells can be reactivated *in vitro* by co-culturing peripheral blood mononuclear cells from previously infected (seropositive) individuals with X-ray irradiated cells of the autologous EB virus-transformed B-cell line at particular responder: stimulator ratios[13,14]. Such preparations are preferentially cytotoxic to the autologous stimulating line and kill allogeneic EB virus-transformed cell lines in an HLA-A and -B antigen-restricted fashion[8–14]. Recent work in this laboratory has shown that cytotoxic T cells of this kind can be established as continuously growing TCGF-dependent cell lines without loss of antigen specificity or of HLA-restriction (L.E.W., M.R., A.B.R. and M.A.E., in preparation).

In the present experiments, EB virus-specific cytotoxic T cells prepared by *in vitro* reactivation from seropositive donor SW (HLA-A11, AW24, B7, BW35) were immediately transferred to TCGF-medium with an equal number of X-ray irradiated cells of the autologous EB virus-transformed cell line SW-EB, as feeders. Foci of proliferating cells appeared within 3 days and the resultant T-cell line, maintained thereafter by twice weekly subculture with feeder cells as above, showed a pattern of EB virus-specific HLA-A and -B antigen-restricted cytotoxicity identical to that of the original reactivated preparation. The results in Fig. 1 show that strong killing of the autologous cell line by TCGF-dependent effector cells was not accompanied by any significant reactivity against EB virus genome-negative human leukaemic cell lines HSB2 and K562, included as indicators of natural killer (NK)-like activity[15]. Furthermore, lysis of allogeneic EB virus-transformed cells was particularly strong when associated with either HLA-A11 (target CM) or BW35 (target EG) antigen matching and progressively weaker with matching through B7 (target JT) and through AW24 (target TH). This pattern of HLA antigen 'preferences' has been observed reproducibly in reactivated cytotoxic preparations from donor SW[14].

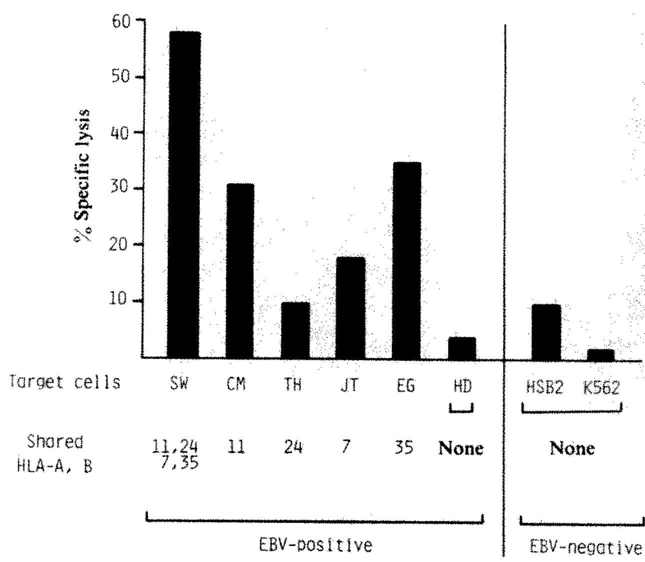


Fig. 1 EB virus-specific, HLA-A and -B antigen-restricted effector function of a TCGF-dependent cytotoxic T-cell line established from donor SW (A11, AW24; B7, BW35). Effector T cells were first obtained by culturing peripheral blood mononuclear cells for 10 days with X-ray irradiated cells of the autologous EB virus-transformed cell line (SW-EB) at a responder:stimulator ratio of 40:1 as described elsewhere[13,14]. Thereafter the T cells were cultured in 2 ml Linbro wells at 2×10^5 cells per well together with 2×10^5 X-ray irradiated SW-EB cells per well in TCGF-medium (RPMI 1640 supplemented with 10% fetal calf serum and 25% TCGF prepared according to the method of Inouyé *et al.*[24]). Cells were collected every 3–4 days and recultured together with fresh stimulator cells in TCGF-medium as above. Thirteen days after their original exposure to TCGF, some of the cells were collected and incubated overnight in TCGF-free medium to remove any residual lectin. Their effector function was tested in a chromium release assay, against the autologous EB virus-transformed cell line, against a range of allogeneic EB virus-transformed cell lines whose degree of HLA-A and -B antigen-matching with the effector cells was as shown, and against EB virus genome-negative cell lines HSB2 and K562 included as indicators of NK-like activity. Results are expressed as percentage specific lysis of ^{51}Cr-labelled target cells in a 5-h assay at an effector: target ratio of 10:1.

0028-0836/82/220413—03$01.00

Table 1 Effect of the monoclonal antibodies W6/32 and Leu 2a on target cell lysis by clone G9

Monoclonal antibody		% Specific lysis of EB virus-transformed cells from donor: SW (A11, AW24; B7, BW35)	BVR (A11, BW35)	EG (A2, A3; BW35, BW40)	SF (A2 A11; B7, BW44)
—	—	29	68	75	2
W6/32	(ref. 21)	2	15	23	13
Leu 2a	(refs 22, 23)	0	15	16	6

Results of a chromium release assay, conducted as described in Fig. 1, using clone G9 effector cells and EB virus-transformed target cells from donors whose HLA-A and -B antigen type was as shown. Levels of percentage specific lysis seen in the standard assay are compared with those seen when target cells were first pretreated for 20 min at 37 °C in the assay wells with a saturating concentration of W6/32 (1:200 final dilution of a purified immunoglobulin preparation) and with those seen when effector cells were first pretreated, again for 20 min at 37 °C in the assay wells, with a saturating concentration of Leu 2a (1:200 final dilution of serum/ascites).

On day 12 following transfer of the effector cell preparation into TCGF-medium, a sample of the cells was cloned at 0.7 cells per well onto a feeder layer of X-ray irradiated SW-EB cells in U-shaped Microtest plate wells. The cultures were maintained with regular supplements of TCGF-medium and, after 2 weeks, those showing active growth were subcultured into replicate U-wells. Some of these replicate cultures provided effector cells for a preliminary screening assay of cytotoxicity against SW-EB and against HSB2 target cells. Of the 55 clones screened in this way, 22 showed preferential lysis of the autologous cell line, and of these, 4 grew sufficiently well thereafter in TCGF-medium to permit extensive analysis of effector function and the establishment of cloned T-cell lines.

Results obtained with effector cells from clone C2 are illustrated in Fig. 2*a*, showing a representative range of targets only, and strongly suggest specificity for the viral antigen LYDMA in association with HLA-A11. In the tests overall, clone C2 killed 8/8 EB virus-transformed cell lines from donors typed as HLA-A11 but did not kill any of 15 corresponding lines from A11-negative donors even though many of these did possess AW24, B7 or BW35 in common with the effector cell donor SW. Moreover there was no lysis of EB virus genome-negative target cells, whether the NK-sensitive cell lines HSB2 and K562 or mitogen-stimulated lymphoblasts from autologous and from A11-matched donors were used. Two further clones, G13 and H8, gave results very similar to those obtained with clone C2. Another clone, G9, when tested against the same range of target cells, gave a different pattern of reactivity (Fig. 2*b*) suggesting specific recognition of LYDMA in association with HLA-BW35. Overall, clone G9 cells killed 9/10 EB virus-transformed target cell lines from donors typed as HLA-BW35 but none of the 11 corresponding lines tested from BW35-negative donors nor any of the relevant EB virus genome-negative target cells.

Previous studies have shown that EB virus-specific cytotoxic T-cell function can be blocked (1) by pretreatment of the target cells with saturating concentrations of monoclonal antibodies (such as W6/32) which bind to a framework determinant common to all HLA-A, -B and -C antigen molecules[16], and (2) by pretreatment of the effector cells with the cytotoxic/suppressor T-cell-specific monoclonal antibody Leu 2a (L.E.W., M.R., A.B.R. and M.A.E., in preparation). The data in Table 1 clearly show that lysis of EB virus-transformed HLA-BW35-bearing

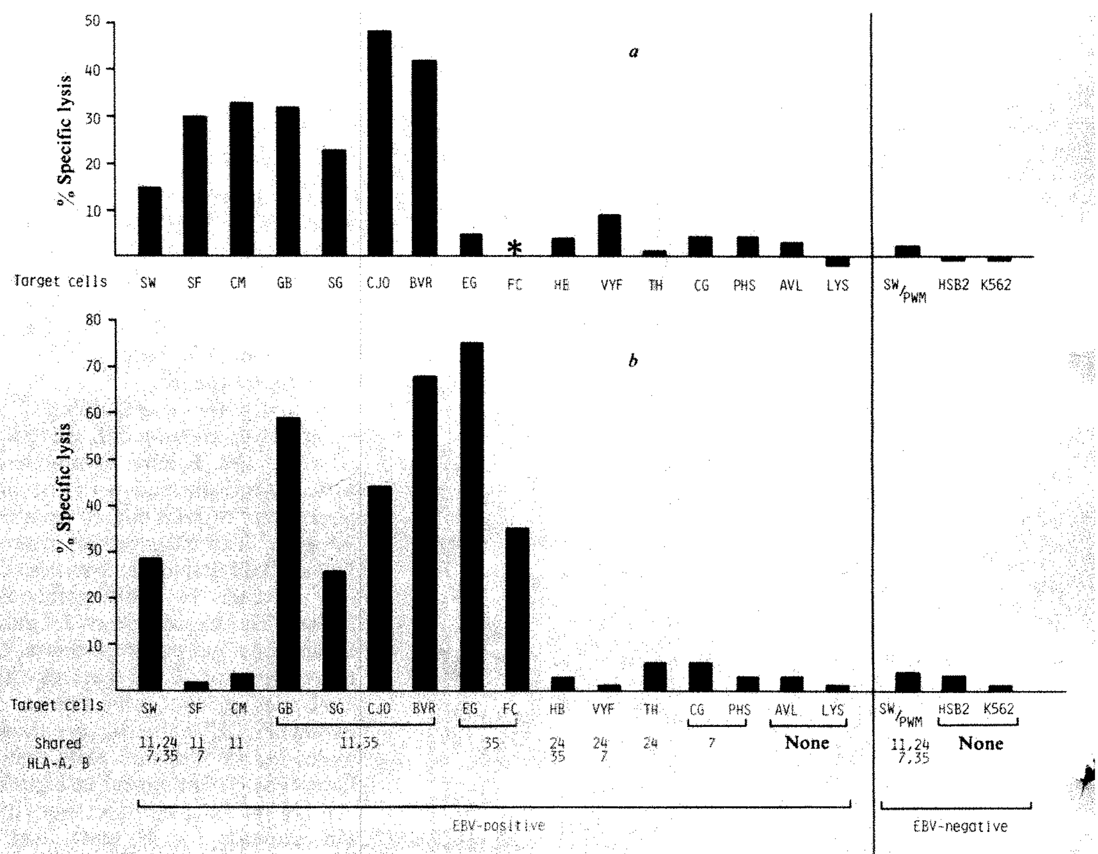


Fig. 2 EB virus-specific effector function of: *a*, clone C2, and *b*, clone G9 derived from the TCGF-dependent cytotoxic T-cell line of donor SW described in Fig. 1. Cloning was performed at an average of 0.7 cells per well seeded into U-shaped Microtest plate wells (Sterilin) containing 2.5×10^3 X-ray irradiated SW-EB cells in TCGF-medium. The cultures were re-fed with fresh TCGF-medium every 3–4 days and those showing active growth were first subcultured after 2 weeks and thereafter expanded into cell lines by regular subculture with X-ray irradiated SW-EB cells in fresh TCGF-medium. The cloned cell lines were tested in chromium release assays, exactly as described in Fig. 1 legend, for cytoxicity against the autologous EB virus-transformed cell line, against a range of allogeneic EB virus-transformed cell lines whose degree of HLA-A and -B antigen-matching with the effector cells was as shown, against autologous pokeweed mitogen (PWM)-stimulated lymphoblasts and against EB virus genome-negative cell lines HSB2 and K562. Results shown for clone C2 are derived from three separate assays performed 35, 38 and 44 days after cloning; results shown for clone G9 are derived from three separate assays performed 31, 50 and 57 days after cloning. * Not tested.

target cells by clone G9 cells is subject to an exactly similar pattern of blocking by these monoclonal antibodies. This adds further support to the view that G9 is indeed a clone of HLA-restricted cytotoxic T cells. In this light, the inability of G9 cells to lyse the cell line from donor HB (HLA-A3, AW24, B14, BW35) is a most interesting result (Fig. 2*b*), particularly as this line did express all four relevant HLA antigens (as defined by serological typing) and was shown in other tests to be sensitive to LYDMA-specific cytolysis by HLA-B14-matched effector T cells (unpublished results). Whilst one could argue that preferential association of LYDMA with HLA-B14 (and possibly with A3 and AW24) on the HB target cell surface might explain the cells' insensitivity to BW35-restricted effector cells, it might alternatively be that donor HB possesses an unusual 'split' of BW35 which effector T cells can distinguish but which is not recognized by current typing sera. Studies with influenza virus-specific human cytotoxic T cells have recently provided examples of this phenomenon within the serologically defined antigens HLA-A2 and A3[17,18].

Clonal analysis of cytotoxic T-cell preparations *in vitro* was first achieved in the H-Y antigen-specific mouse model system[19] and has more recently been investigated using this same antigen-specific system in man[20]. The present report describes the first successful application of this technique in studying the human cytotoxic T-cell response to a natural pathogen. The results show that the response to EB virus is indeed polyclonal with individual reactivities recognizing the viral target antigen in the context of one particular HLA determinant. Since EB virus-specific effector T-cell preparations can be made from most previously infected individuals, it should be possible to establish a library of cytotoxic clones to cover the full range of HLA-A and -B antigen restrictions. Monospecific reagents of this kind could not only facilitate the biochemical analysis of cytotoxic T-cell function but also prove even more valuable as probes for HLA-A and -B antigen polymorphism.

We thank Elizabeth Gough, Ann Green and Jill Sweet for technical help and Dr B. A. Bradley and staff of the UK Transplant Services, SW Regional Transfusion Centre, Bristol, for HLA-A, -B and -C antigen typing of the cell donors. The gifts of monoclonal antibodies W6/32 and Leu 2a from Dr M. J. Crumpton, Imperial Cancer Research Fund Laboratories, and Dr R. B. Evans, Sloan-Kettering Institute for Cancer Research, respectively, are gratefully acknowledged. This work was supported by the MRC (grant G979/1123) and assisted by the Cancer Research Campaign, London, out of funds donated by the Bradbury Investment Corporation of Hong Kong.

Received 1 February; accepted 6 April 1982.

1. Zinkernagel, R. M. & Doherty, P. C. *Nature* **251,** 547–548 (1974).
2. Doherty, P. C. & Zinkernagel, R. M. *J. exp. Med.* **141,** 502–512 (1975).
3. Doherty, P. C., Blanden, R. V. & Zinkernagel, R. M. *Transplantn Rev.* **29,** 89–124 (1976).
4. McMichael, A. J., Ting, A., Zweerink, H. J. & Askonas, B. A. *Nature* **270,** 524–526 (1977).
5. Biddison, W. E. & Shaw, S. *J. Immun.* **122,** 1705–1709 (1979).
6. Shaw, S. & Biddison, W. E. *J. exp. Med.* **149,** 565–575 (1979).
7. McMichael, A. J., Parham, P., Brodsky, F. M. & Pilch, J. R. *J. exp. Med.* **152,** 195s–203s (1980).
8. Rickinson, A. B., Wallace, L. E. & Epstein, M. A. *Nature* **283,** 865–867 (1980).
9. Misko, I. S., Moss, D. J. & Pope, J. H. *Proc. natn. Acad. Sci. U.S.A.* **77,** 4247–4250 (1980).
10. Moss, D. J., Wallace, L. E., Rickinson, A. B. & Epstein, M. A. *Eur. J. Immun.* **11,** 686–693 (1981).
11. Fukukawa, T. *et al. J. Immun.* **126,** 1697–1701 (1981).
12. Tsoukas, C. D. *et al. J. Immun.* **126,** 1742–1746 (1981).
13. Rickinson, A. B. *et al. Cancer Res.* **41,** 4216–4221 (1981).
14. Wallace, L. E. *et al. Cell. Immun.* **67,** 129–140 (1982).
15. Jondal, M. & Targan, S. *J. exp. Med.* **147,** 1621–1636 (1978).
16. Wallace, L. E., Moss, D. J., Rickinson, A. B., McMichael, A. J. & Epstein, M. A. *Eur. J. Immun.* **11,** 694–699 (1981).
17. Biddison, W. E., Ward, F. E., Shearer, G. M. & Shaw, S. *J. Immun.* **124,** 548–552 (1980).
18. Biddison, W. E., Shearer, G. M. & Shaw, S. *J. Immun.* **127,** 2231–2235 (1981).
19. von Boehmer, H. *et al. Eur. J. Immun.* **9,** 592–597 (1979).
20. Goulmy, E. *et al. J. exp. Med.* **152,** 182s–190s (1980).
21. Barnstable, C. J. *et al. Cell* **14,** 9–20 (1978).
22. Ledbetter, J. A. *et al. J. exp. Med.* **153,** 310–323 (1981).
23. Evans, R. L. *et al. Proc. natn. Acad. Sci. U.S.A.* **78,** 544–548 (1981).
24. Inouye, H., Hank, J. A., Alter, B. J. & Bach, F. J. *Scand. J. Immun.* **12,** 149–154 (1980).

Product of a transferred $H\text{-}2L^d$ gene acts as restriction element for LCMV-specific killer T cells

Anders Örn, Robert S. Goodenow & Leroy Hood

Division of Biology, California Institute of Technology, Pasadena, California 91125, USA

Peter R. Brayton*, Jerold G. Woodward†, Richard C. Harmon† & Jeffrey A. Frelinger†

Departments of *Neurology and †Microbiology, University of Southern California Medical School, Los Angeles, California 90033, USA

The major histocompatibility complex (MHC) of the mouse, *H-2* on chromosome 17, contains several families of genes encoding cell-surface molecules which have a role in mediating immune responses[1,2]. The class I genes encode a family of homologous membrane proteins including the transplantation antigens K, D and L. These antigens reflect extensive genetic polymorphism which is apparent in the many different class I gene constellations or haplotypes found in mice (for example, BALB/c mice exhibit the $H\text{-}2^d$ haplotype and their class I molecules are denoted K^d, D^D and L^d). Transplantation antigens serve as targets for T-cell killing in allogeneic immune responses such as *in vivo* graft rejection[3] and destruction of allogeneic cells by cytotoxic T cells *in vitro*[4]. However, the physiological role of transplantation antigens may be to serve as restricting elements in virus-mediated T-cell killing of infected self cells. Virus infection of mice generates killer T cells whose receptors must interact with the foreign viral antigen and a class I molecule or restricting element for the cytotoxic effector function to be activated[5,6]. Thus the T-cell receptor recognizes the viral antigen in the context of a class I molecule. To study the interaction between the T-cell receptor and the class I restricting element, we have used the mouse L-cell transformant 8-5 which expresses L^d molecules[7] and the K7-65 transformant expressing K^d molecules (R.S.G. *et al.*, in preparation). Mouse L cells are fibroblasts derived from C3H mice of $H\text{-}2^k$ haplotype and monoclonal antibodies can be used to distinguish $H\text{-}2^d$ molecules from the endogenous $H\text{-}2^k$ products. Recently, we have demonstrated that both L^d (ref. 8) and K^d molecules (unpublished data) expressed on transformed L cells can act as target antigens for alloreactive cytotoxic T lymphocytes. Here we show that the L^d molecule on transformed mouse L cells can serve as a restricting element in lymphocytic choriomeningitis virus (LCMV) infection, whereas its K^d counterpart cannot.

Recently, we described (ref. 7 and R.S.G. *et al.*, in preparation) the production of L-cell transformants expressing L^d and K^d molecules. Briefly, thymidine kinase-negative (tk^-) L cells were co-transformed with the herpesvirus thymidine kinase gene alone or with the *tk* gene together with a cloned L^d or K^d gene. The transformed cells were selected for thymidine kinase (TK) activity and screened for the expression of L^d or K^d molecules by radioimmunoassay with monoclonal antibodies. The expression of L^d and K^d was confirmed by two-dimensional gel electrophoresis. In this manner we generated three types of tk^+ target cell—those expressing K^k and D^k molecules (L-tk^+ cells), those expressing K^k, D^k and K^d molecules (K7-65 cells) and those expressing K^k, D^k and L^d molecules (8-5 clone). Note that mice (or cells) of the $H\text{-}2^k$ haplotype do not appear to express L antigens.

C3H and BALB/c mice were infected with LCMV to generate cytotoxic T cells, and spleen cells from immunized mice were assayed in a 4-h ^{51}Cr-release assay against various target cells (Fig. 1). The cytotoxic T cells generated in C3H mice killed the LCMV-infected target cells from cloned 8-5 cells ($K^kD^kL^d$), L-tk^+ cells (K^kD^k), and K7-65 cells ($K^kD^kK^d$) (Fig. 1*a*). The same cytotoxic T cells killed neither LCMV-infected

0028–0836/82/220415—03$01.00

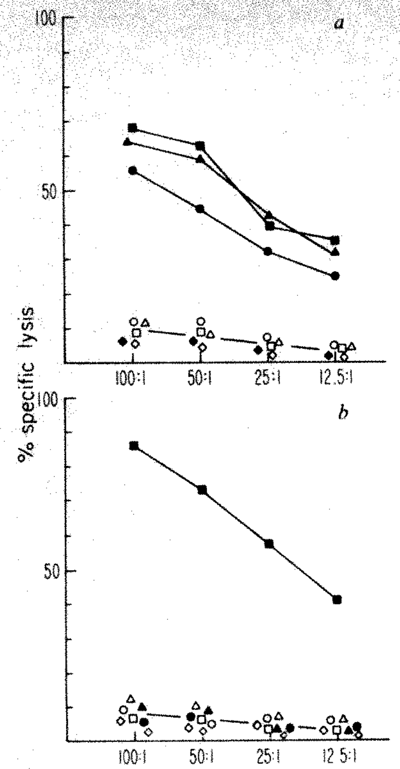


Fig. 1 Killing of LCMV-infected L-cell transformants by spleen cells from LCMV-infected C3H (H–2^k) and BALB/c (H–2^d) mice. Groups of two 6–8-week-old C3H/Bang and BALB/c mice were infected intraperitoneally with 2.5×10^2 plaque-forming units of the Armstrong strain of LCMV 7 days before cytotoxic assay. Spleens were removed and single cell suspensions prepared. L-cell transformants were grown in hypoxanthine–aminopterin–thymidine medium and Nulli-2A teratocarcinoma cells in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum (FCS). The presence of L^d and K^d molecules on the transformants was monitored by radioimmunoassay using monoclonal antibodies as described previously[7]. Target cells were infected with LCMV 40 h before cytotoxicity assay at a multiplicity of infection of 0.5, then trypsinized, washed, and labelled for 1 h at 37 °C with 200 μCi ^{51}Cr in 0.5 ml RPMI supplemented with 10% FCS and 20 mM HEPES. After washing, 10^4 target cells were mixed in triplicates in microtitre plates with varying numbers of effector spleen cells in a total volume of 200 μl. Plates were incubated at 37 °C, 5% CO_2 for 4 h, then centrifuged and 100 μl of the supernatant were removed and counted in a γ-counter. Per cent specific lysis was calculated according to the formula: $\text{c.p.m.}_{\text{exp}} - \text{c.p.m.}_{\text{spon}}/\text{c.p.m.}_{\text{max}} - \text{c.p.m.}_{\text{spon}} \times 100$, where $\text{c.p.m.}_{\text{spon}}$ (spontaneous) and $\text{c.p.m.}_{\text{max}}$ (maximum) were determined in cultures containing medium or detergent, respectively; $\text{c.p.m.}_{\text{exp}}$, c.p.m. (experimental). The s.e.m. never exceeded 5%. The figure shows results for spleen effector cells from *a*, LCMV-infected C3H/Bang mice, and *b*, LCMV-infected BALB/c mice. Target cells: ■, 8-5 (L^d)–LCMV; □, 8-5, uninfected; ▲, L-tk^+–LCMV; △, L-tk^+, uninfected; ●, K7-65 (K^d)–LCMV; ○, K7-65, uninfected; ◆, Nulli-2A–LCMV; ◇, Nulli-2A, uninfected.

Nulli-2A teratocarcinoma cells, which do not express class I molecules[9], nor uninfected target cells. These data demonstrate that LCMV-specific T cells from C3H mice require the viral antigen and a K^k or D^k restricting element. In contrast, LCMV-restricted cytotoxic T cells from BALB/c mice killed LCMV-infected 8-5 cells but not any of the other target cells (Fig. 1*b*). These data demonstrate that BALB/c mice can use the L^d but not the K^d molecule as a T-cell restricting element in response to LCMV. This is consistent with reports on other *H-2*-restricted systems which claim that L^d can act as a restricting element[10–12].

We then raised cytotoxic lymphocytes to LCMV in C-H–2^{dm2} mice. These are mutant mice of the H–2^d haplotype which fail to express an L^d molecule[13,14]. Cytotoxic lymphocytes from the *dm2* mice lysed neither LCMV-infected 8-5 ($K^kD^kL^d$) nor L-tk^+ (K^kD^k) cells (Fig. 2*a*). Control experiments with LCMV-induced cytotoxic T cells in C3H mice again demonstrated lysis of both the 8-5 and L-tk^+ cells as expected (Fig. 2*c*). A similar analysis with LCMV-specific killer T cells raised in B10.Q mice of the H–2^q haplotype lysed neither the 8-5 nor L-tk^+ cells (Fig. 2*d*). This control experiment is interesting because mouse strains of the H–2^q haplotype are known to express an L^q antigen serologically related to the L^d product[15–17]. Thus, the T-cell receptor seems to interact with that portion of the L^d antigen which distinguishes it from its L^q counterpart.

We further tested the specificity of the L^d restricting element in LCMV-infected 8-5 cells by inhibiting the T-cell lysis of these cells with specific antibodies. Two L^d-specific alloantisera were raised by immunizing C3H mice with 8-5 cells as described in detail elsewhere[18], and by immunizing C-H–2^{dm2} mice with BALB/c spleen cells. In addition, we used two monoclonal anti-L^d antibodies (28-14-8S and 4-9.4), a monoclonal anti-D^d antibody (34-2-12), a monoclonal K^d antibody (6-27.10) and normal mouse serum. The results of these experiments are shown in Fig. 3*a*, *b*. All antibodies directed against the L^d molecule inhibited the LCMV-restricted cytotoxic T-cell killing of the L^d-transformed 8-5 cells. In contrast, normal mouse serum as well as monoclonal antibodies to the D^d and K^d antigens, did not inhibit this cytotoxic T-cell reaction. These data again suggest that the L^d molecule serves as a typical restricting element in LCMV-induced T-cell killing.

Thus, the gene transfer procedure results in the expression of class I molecules which are inserted in the membrane of a transformed cell in a normal manner so that they can act as restricting elements for virus-induced cytotoxic T cells. An analysis of various viral infections using the different transplantation antigens should allow us to delineate the roles of the class I molecules in specific virus infections. Moreover, this system, together with *in vitro* mutagenesis, should allow us to elucidate the nature of the interactions between the T-cell receptor and class I molecules. The major advantage of *in vitro* mutagenesis is the ability to generate mutants that carry alterations in specific regions of the class I molecules. In contrast, conventional mutants are selected only for altered T-cell recognition. Thus, our ability to transfer normal and altered class

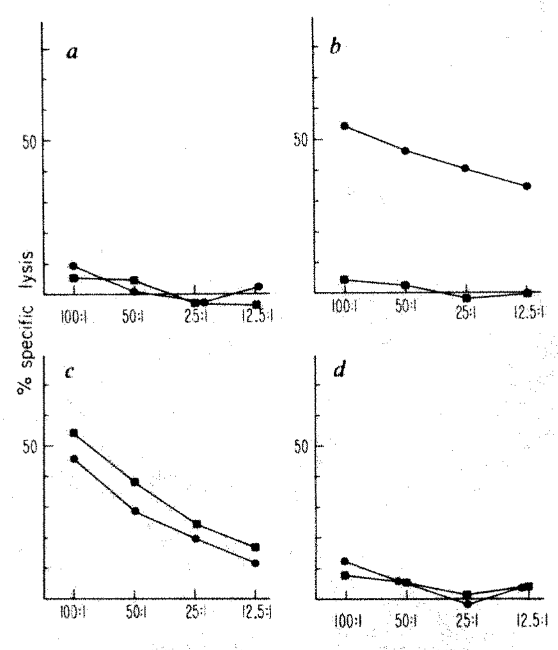


Fig. 2 Killing of LCMV-infected L-cell transformants by spleen cells from different LCMV-infected mouse strains. For experimental details see Fig. 1 legend. Spleen effector cells from LCMV-infected mice: *a*, C-H–2^{dm2} (H–2^d mutant); *b*, BALB/c (H–2^d); *c*, C3H/Bang (H–2^k); *d*, B10.Q (H–2^q). Target cells: ●, 8-5 (L^d)–LCMV; ■, L-tk^+–LCMV. Uninfected target cells never showed >10% lysis.

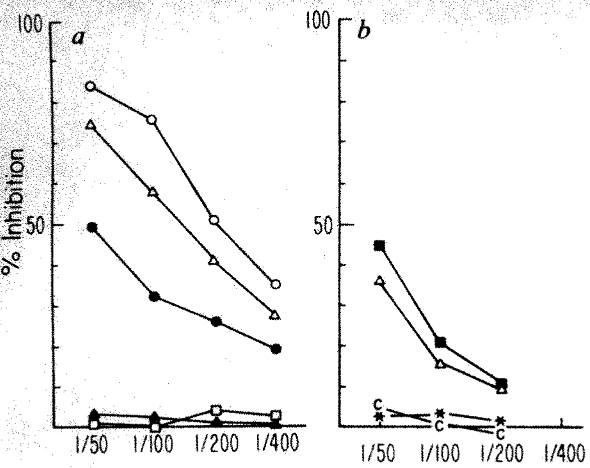


Fig. 3 Inhibition of *H-2*-restricted LCMV-specific cytotoxicity by antibodies directed against different *H-2* antigens. Various antisera or monoclonal reagents were added at the start of the Cr-release assay in the dilutions shown. The cytotoxic reaction inhibited was that of LCMV-infected BALB/c spleen cells against LCMV-infected 8-5 cells. Per cent inhibition was calculated as follows:

$$1 - \frac{\%\ \text{specific lysis in the presence of antibody}}{\%\ \text{specific lysis in the absence of antibody}} \times 100$$

a and *b* show two separate experiments: *a*, 30.5% lysis in the absence of blocking antibody at effector/target ratio of 50:1. Blocking antibodies: ○, C3H/Bang anti-8-5 alloantiserum; △, 28-14-8S (anti-L^d monoclonal antibody); ●, 4-9.4 (anti-L^d monoclonal antibody); ▲, 34-2-12 (anti-D^d monoclonal antibody); and □, normal mouse serum. The specificities of these monoclonal antibodies are described elsewhere[8,19]. *b*, 41.0% Lysis in the absence of blocking antibody at effector/target ratio of 100:1. Blocking antibodies: ■, C-*H-2*^{dm2} anti-2BALB/c alloantiserum; △, 28-14-8S (anti-L^d monoclonal antibody), C, 6-27.10 (anti-K^d monoclonal antibody); and *, normal mouse serum.

I genes into foreign cells provides us with one of the first opportunities to elucidate the molecular basis of cell–cell interactions.

This work was supported by NIH grants GM 06965, CA 22662, CA 26199 and CA 25911. We thank Drs David Sachs and Ted Hansen for monoclonal antibodies and antisera; Dr Roger Melvold for supplying the C-*H-2*^{dm2} mice and Dr Peter Stern for the Nulli-2A cell line. R.S.G. is a Senior Lievre Fellow of the California Division of the American Cancer Society. J.G.W. is a recipient of a National Arthritis Foundation Postdoctoral Fellowship. R.C.H. and P.R.B. are supported by NRSA-NS-07149 and J.A.F. is the recipient of an American Cancer Society Faculty Research Award.

Received 23 March; accepted 14 April 1982.

1. Klein, J. *Science* **203**, 516–521 (1979).
2. Snell, G. D. *Science* **213**, 172–178 (1981).
3. Snell, G. D., Dausset, J. & Nathenson, S. (eds) *Histocompatibility* (Academic, New York, 1976).
4. Alter, B. J. *et al. J. exp. Med.* **137**, 1303–1309 (1973).
5. Doherty, P. C. & Zinkernagel, R. M. *J. exp. Med.* **141**, 502–507 (1975).
6. Zinkernagel, R. M. & Doherty, P. C. *Adv. Immun.* **27**, 51–177 (1979).
7. Goodenow, R. S. *et al. Science* **215**, 677–679 (1982).
8. Woodward, J. G. *et al. Proc. natn. Acad. Sci. U.S.A.* (in the press).
9. Stern, P., Martin, G. R. & Evans, M. J. *Cell* **6**, 455–465 (1975).
10. Biddison, W. E., Hansen, T. H., Levy, R. B. & Doherty, P. C. *J. exp. Med.* **148**, 1678–1686 (1978).
11. Blanden, R. V., McKenzie, I. F. C., Kees, U., Melvold, R. W. & Kohn, H. I. *J. exp. Med.* **146**, 869–880 (1977).
12. Levy, R. B., Shearer, G. M. & Hansen, T. H. *J. Immun.* **121**, 2263–2269 (1978).
13. McKenzie, I. F. C., Morgan, G. M., Melvold, R. W. & Kohn, H. I. *Immunogenetics* **4**, 333–347 (1977).
14. Hansen, T. H. *et al. J. Immun.* **126**, 1713–1716 (1981).
15. Hansen, T. H. & Sachs, D. H. *J. Immun.* **121**, 1469–1472 (1978).
16. Huang, C-M., Huang, S. & Klein, J. *Immunogenetics* **9**, 173–182 (1979).
17. Ozato, K., Hansen, T. H. & Sachs, D. H. *J. Immun.* **125**, 2473–2477 (1980).
18. Woodward, J. G. *et al.* in *B and T Cell Tumors: Biological and Clinical Aspects* (eds Vitetta, E. & Fox, C. F.) (Academic, New York, in the press).
19. Ozato, K., Meyer, R. & Sachs, D. H. *Transplantation* (in the press).

Sensitivity and resistance of human tumour cells to interferon and $rI_n \cdot rC_n$

Stanley L. Lin*, James J. Greene & P. O. P. Ts'o

The Johns Hopkins University, School of Hygiene and Public Health, Division of Biophysics, 615 North Wolfe Street, Baltimore, Maryland 21205, USA

W. A. Carter

Herbert L. Orlowitz Institute, Hahnemann Medical College and Hospital, Philadelphia, Pennsylvania 19102, USA

Interferon exerts, in addition to its well recognized antiviral activity, an antiproliferative effect[1,2] and has been found to reduce plasminogen activator secretion[3], to re-establish contact inhibition[4] and to inhibit the growth in semi-solid agar[5] of neoplastic cells. Double-stranded RNA (dsRNA) is both an interferon inducer and an activator of two interferon-induced enzymes, a protein kinase and a 2',5' oligoadenylate synthetase[6]. However, the molecular bases for the antiproliferative and antineoplastic effects of interferon, and their relationship to dsRNA, are unknown. We have now characterized the antiproliferative effects of interferon and dsRNA on the human HT1080 fibrosarcoma[7] and RT4 epidermal carcinoma[8] cell lines. Luria–Delbrück fluctuation analysis[9] of interferon resistance indicates that the resistance trait is a non-randomly acquired phenotype which was probably induced by interferon at the time of selection. The interferon-resistant (IFN^r) cells retained sensitivity to the antiviral effects of interferon, as well as to the antiproliferative effects of dsRNA.

Clonal populations of interferon-sensitive (IFN^s) cells were isolated from both HT1080 and RT4 cell lines. The HT1080 (HT1080-2) and RT4 (RT4-1) clonal lines demonstrated a complete growth inhibition at 400 and 80 IU ml^{-1} human fibroblast interferon (HuIFN-β) respectively, when examined over a 72 h period (data not shown). The fluctuation analysis was initiated by establishing 10 independent cultures from each original colony and continuously growing each culture for 16 population doublings, as described in Table 1 legend, before plating in selective medium. Results are shown in Table 1. In all cultures of the HT1080-2 population, 3.7–4.5% of the cells were capable of colony formation in interferon-containing medium (1,000 IU ml^{-1}). There was no difference in average colony size or cell morphology between control and interferon-treated cells. The frequency distribution of the number of resistant colonies obtained from this series of sister cultures exhibited a variance of 7.73. The variance of the distribution of the number of colonies in a large single culture (v_2, random sampling control) was 2.04. After correction for the random sampling control, the variance-to-mean ratio [v.m.r., $(v_1 - v_2)$/mean] of the frequency distribution of the number of resistant colonies obtained from the series of cultures was 0.15. In contrast, when RT4-1 cells were similarly treated, interferon-resistant colonies arose at a frequency of 0.11%, 30–40-fold less than the frequency observed for HT1080-2 cells (Table 1). However, the v.m.r. of the distribution obtained from the series of 10 sister cultures was low—0.31. Simultaneous examination of the HT1080-2 cultures for acquisition of 6-thioguanine resistance (6TG^r), an acquired trait of known spontaneous origin, resulted in a corrected v.m.r. of 6.79 (Table 1). The high v.m.r. is indicative of a randomly acquired trait and indicates that 16 population doublings were sufficient to distinguish between randomly and non-randomly acquired phenotypes. Using the p_0 method of Luria and Delbrück[9], a mutation rate of 1.8×10^{-8} mutants per cell per generation can be calculated from the values presented in Table 1 for the acquisition of 6TG^r. This rate is consistent with known mutation rates obtained for various loci in other systems[10], and thus verifies our procedure.

* Present address: The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA.

Examination of isolated colonies derived from resistant populations revealed that IFN^r cells possessed the same population doubling times in the presence and absence of interferon at concentrations normally resulting in complete cessation of growth (Fig. 1*a*, *c*). However, consistent with the notion that IFN^r is an induced trait, resistant colonies reverted to a sensitive state, albeit after varying lengths of time. One clone of HT1080-IFN^r cells retained the resistance trait after 4 weeks growth in interferon-free medium and reverted to a sensitive state by 15 weeks after removal of interferon from the culture medium (Fig. 1*a*), whereas a clone of resistant RT4 cells was fully resistant to interferon immediately after isolation, but reverted completely by 4 weeks after removal of interferon from the culture medium (Fig. 1*c*). When used as assay cells to titre an identical solution of interferon by the dye uptake method of Finter[11], resistant cells during their resistant state were at least as sensitive as their parental populations to the antiviral effects of interferon when vesicular stomatitis virus was used as the challenge virus (Table 2), suggesting that resistance to the antiproliferative effects of interferon do not necessarily correlate with changes in sensitivity to its antiviral effects. It should be mentioned, however, that different viruses can be affected by different mechanisms pertaining to the antiviral state[12–14].

In view of the participation of dsRNAs in the establishment of the antiviral state and the toxicity of exogenous dsRNA[15], we characterized the sensitivities of the interferon-resistant lines towards the antiproliferative effects of the synthetic dsRNA, $rI_n \cdot rC_n$ (Fig. 1). Resistant lines of both HT1080-2 (Fig. 1*b*) and RT4-1 (Fig. 1*d*) displayed the same sensitivity to

Table 1 Luria-Delbrück fluctuation analysis for the acquisition of interferon resistance

	HT1080		RT4
Culture	IFN^r colonies per 10^3 cells	6TG^r colonies per 10^6 cells	IFN^r colonies per 10^5 cells
1	39.3	7.6	126
2	36.9	2.4	145
3	38.2	6.1	127
4	35.7	14.2	127
5	41.1	0.0	97
6	42.5	0.0	102
7	44.6	18.0	96
8	39.1	0.0	119
9	37.1	2.5	122
10	35.9	2.7	108
Variance	7.73	35.25	224.09
Mean	39.04	5.35	116.90
Random sampling variance	2.04	0.16	187.40
Corrected variance/mean	0.15	6.56	0.31

HT1080 cells were cloned in semi-solid agar as described by Laug *et al.*[21]. A vigorously growing colony was isolated and transferred to a 35 mm dish containing 2 ml growth medium, disbursed by repeated pipetting and allowed to attach and grow. The resultant monolayer was detached from the plate by trypsinization and was transferred to a 75 cm^2 flask for serial propagation. Cells comprising the resultant confluent monolayer were detached by trypsinization and 10 independent cultures were initiated by inoculating 10 35 mm dishes at 100 cells per dish. On reaching confluence, the total number of cells from each culture was detached by trypsinization and transferred to a separate 75 cm^2 flask to maintain the 10 cultures independently. After a total of 16 population doublings from initial establishment of the independent cultures, confluent monolayers in 75 cm^2 flasks were detached by trypsinization and the frequency of IFN^r cells in each independent culture determined by placing cells in liquid medium containing HuIFN-β at a plating density of 1×10^4 cells per 100mm dish. Fluctuation analysis for the acquisition of 6TG^r was simultaneously performed using selective conditions described previously[22]. An RT4 clone was isolated by the agar overlay method described in ref. 23, transferred to a 35 mm dish containing 2 ml growth medium and then handled in identical fashion to the HT1080 clone. RT4 cells were plated at a density of 1×10^5 cells per 100 mm dish in interferon-containing medium. Selective medium consisted of growth medium containing HuIFN-β HuIFN-β at a plating density of 1×10^4 cells per 100 mm dish. Fluctuation analysis or 6TG (3 μg ml^{-1}). Results shown are corrected for plating efficiency. Random sampling variance was determined by analysing the frequency distribution observed in a single large culture in the same way as for independent cultures. Correction for sampling error was performed by subtracting the random sampling variance from the variance of the independent cultures before division by the mean. IFN-β (fibroblast-derived) used in these studies had specific activities of 2×10^6 (HEM Research, Rockville, Maryland) to 2×10^7 (Dr W. A. Carter) IU per mg protein after purification by sequential concanavalin A–Sepharose→phenyl-Sepharose affinity chromatographies[24].

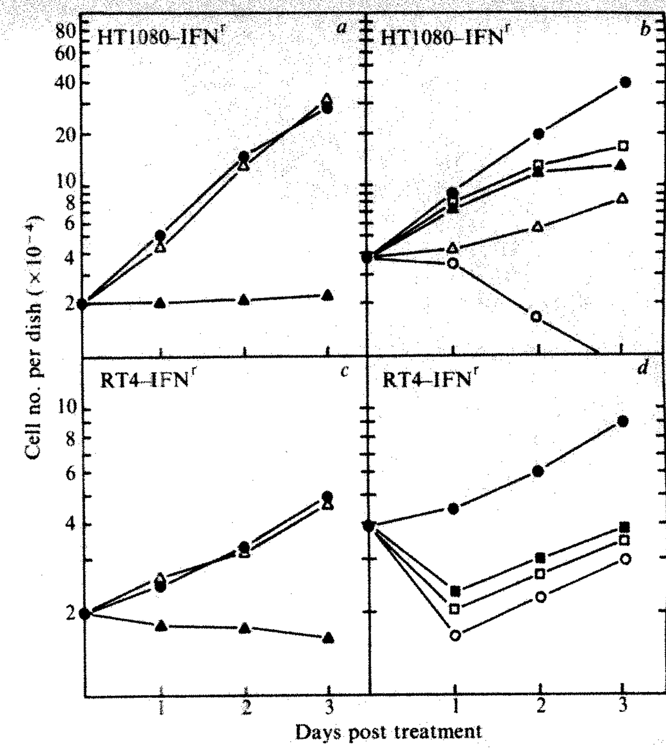


Fig. 1 Heritability of resistance to the growth inhibitory effects of interferon, and effects of $rI_n \cdot rC_n$ on the growth of IFN^r cells in culture. Colonies growing in interferon-containing medium were isolated and grown in the presence of medium containing 1,000 (HT1080) or 100 (RT4) IU ml^{-1} HuIFN-β. After ~16 population doublings, HuIFn-β was removed from the culture medium and cells were examined for interferon resistance at various passages after growth in interferon-free medium (*a*, *c*) or for sensitivity to $rIn \cdot rC_n$-mediated toxicity. (*a*) Growth of HT1080-IFN^r cells in the absence (●) or presence of HuIFN-β (1,000 IU ml^{-1}) after culture for 4 (△) and 15 weeks (▲) in interferon-free medium; (*c*), growth of RT4-IFN^r cells in the absence (●) or presence of HuIFN-β (100 IU ml^{-1}) after culture for 0 (△) and 4 (▲) weeks in interferon-free medium. For measuring dsRNA-mediated toxicity, 5×10^4 (HT1080-IFN^r)or 1×10^5 (RT4-IFN^r) cells were plated out per 35 mm dish and allowed to attach overnight. The following day (time 0), cell number per dish was determined, remaining monolayers were refed with various concentrations of $rI_n \cdot rC_n$ and cell number per dish was determined every 24 h thereafter. (*b*), HT1080-IFN^r cells grown in 0 (●), 0.1 (□), 0.2 (▲), 0.5 (△) and 2.5 (○) mM $rI_n \cdot rC_n$; (*d*), RT4-1IFN^r cells grown in 0 (●), 0.1 (■), 5.0 (□) and 10.0 (○) μM $rI_n \cdot rC_n$.

$rI_n \cdot rC_n$-mediated toxicity as did their respective parental HT1080-2 and RT4-1 lines (not shown). For HT1080-IFN^r populations, 0.5–1.0 mM $rI_n \cdot rC_n$ was required to inhibit cell proliferation completely, with $rI_n \cdot rC_n$ concentrations >1 mM producing marked cell death within 24 h. In these conditions, interferon concentrations secreted in the medium did not exceed 3 IU ml^{-1}. Detectable $rI_n \cdot rC_n$-induced toxicity in RT4-IFN^r cells was observed at concentrations as low as 0.1 μM, with no induced interferon being detectable in the medium. However, in contrast to the prolonged cytotoxic effect observed in HT1080-2 cells, some clones of RT4-IFN^r cells could recover from dsRNA-mediated toxicity after 24 h and exhibit the same growth rate as before treatment. This recovery was not the result of rapid degradation of $rI_n \cdot rC_n$ as re-addition of $rI_n \cdot rC_n$ to the culture medium at 24 or 48 h did not alter the recovered growth rate (data not shown). Addition of antisera, directed against HuIFN-β (final concentration 100 neutralizing units per ml), to the culture medium of $rI_n \cdot rC_n$-treated cells did not reduce or otherwise alter the extent of cytotoxicity induced by $rI_n \cdot rC_n$ (not shown), suggesting that low interferon concentration was not involved in $rI_n \cdot rC_n$-mediated toxicity.

The Luria–Delbrück fluctuation analysis, originally developed for microbial systems, provides a statistical method whereby the acquisition of a certain phenotype by a population

Table 2 Sensitivity of IFN^s and IFN^r lines to the antiviral effects of interferon

Cell line	Phenotype	Titre
HT1080-2	IFN^s	1.5
HT1080-2	IFN^r	4.5
RT4-1	IFN^s	30.2
RT4-4	IFN^r	28.0

Confluent monolayers (1.5 cm diameter) of IFN^S or IFN^r cells from both HT1080 and RT4 lines were used as assay cells for the titration of a stock solution of interferon (concentration 50 IU ml^{-1}), using the dye uptake method of Finter[11] with vesicular stomatitis virus as the challenge virus.

of cells can be determined to be random (spontaneous) or non-random (induced)[9]. When a series of separate cultures, all initiated from a single clone, is allowed to undergo a suitable number of population doublings, the variation of the frequency of resistant colonies obtained in each culture will be large if the resistant trait is randomly acquired, or small if the trait is non-randomly acquired. Theoretically, the v.m.r. will be ≤ 1 if the trait is acquired non-randomly, or $\gg 1$ if the trait is acquired randomly. The acquisition of interferon resistance when analysed in this way, provided v.m.rs of <1, suggesting that interferon resistance does not result from a spontaneous event, such as mutation or chromosomal rearrangement, but rather is a phenotype induced by interferon treatment. This conclusion is supported by the demonstration reported here, and elsewhere[16], that resistant cells, previously sensitive to the growth inhibitory effects of interferon, revert to a sensitive state. As no difference in average size was observed for colonies formed in control and interferon-containing medium, induction of resistance must occur rapidly, within 24–48 h after interferon exposure. Alternatively, cells in the population might exist in either an IFN^r or IFN^s state, depending on their physiological or developmental status. Addition of interferon would both select for the subpopulation of pre-existing resistant cells and maintain the resistant trait.

Our results contrast with those of Gresser *et al.*[17], who by similar analysis concluded that the resistance phenotype was randomly acquired and appeared at a frequency consistent with somatic mutation. However, Gresser *et al.* obtained their high v.m.r. from comparison of resistance displayed by different clones isolated from a mass tumour cell culture. Recent observations that clonal isolates from a population of tumour cells can vary widely in sensitivity to the antiproliferative effects of interferon[16] indicate that such interclonal comparison cannot be used for determining the randomness of acquisition of interferon resistance. Moreover, the resistant cells we used, as opposed to those described by Gresser *et al.*[17], retained their sensitivity to the antiviral effects of interferon, indicating that resistance was not due simply to the absence of an interferon receptor, and is consistent with previous reports where the antiviral state could be induced by interferon in Rous sarcoma virus-transformed human fibroblasts resistant to the growth inhibitory effects of interferon[18].

Resistant cells retained a sensitivity towards dsRNA-induced toxicity that did not differ from that exhibited by the IFN^s parental cell lines. This is an important consideration in the development of chemotherapeutic regimens for the treatment of cancer with interferon and it suggests that interferon used in combination with dsRNA would be much more effective in the treatment of cancer than interferon alone. In addition, the data argue against a role for the known interferon-induced, dsRNA-dependent enzymes in the antiproliferative effects of interferon. No reduction in interferon-mediated growth inhibition was observed in mitogen-stimulated normal human fibroblasts pretreated with interferon for periods of up to 72 h (ref. 19) and continuous subculture of a sensitive line of normal human fibroblasts in the presence of HuIFN-β for up to 6 weeks did not result in the appearance of a resistant population[20], suggesting that normal cells do not readily acquire interferon resistance. This difference between the ability of normal and neoplastic cells to acquire resistance to the growth inhibitory effects of interferon might be related to the mechanisms of uncontrolled growth characteristic of tumour cells. Further delineation of interferon-mediated growth inhibition could be important in elucidating a fundamental difference between normal and neoplastic cells, as well as in establishing the molecular basis for the antiproliferative effects of interferon.

Received 11 January; accepted 5 April 1982.

1. Stewart, W .E. II, Gresser, I., Tovey, M., Bandu, M. T. & Le Goff, S. *Nature* **262**, 300–302 (1976).
2. Pfeffer, L. M., Murphy, J. S. & Tamm, I. *Expl Cell Res.* **121**, 111–120 (1979).
3. Schroder, E., Chou, I.-N., Jaken, S. & Black, P. H. *Nature* **276**, 828–829 (1978).
4. Knight, E. Jr *J. Cell Biol.* **56**, 846–849 (1973).
5. Gresser, I., Thomas, M.-T. & Brouty-Boyé, D. *Nature* **231**, 20–21 (1971).
6. Farrel, P. J. *et al. Proc. natn. Acad. Sci.* **75**, 5893–5897 (1978).
7. Rasheed, S., Nelson-Rees, W. A. Toth, E. M., Arnstein, P. & Gardner, M. B. *Cancer* **33**, 1027–1033 (1974).
8. Rigby, C. C. & Franks, L. M. *Br. J. Cancer* **24**, 746–754 (1970).
9. Luria, S. E. & Delbrück, M. *Genetics* **28**, 491–511 (1943).
10. Barrett, J. C., Crawford, B. D. & Ts'o, P. O. P. in *Advances in Modern Environmental Toxicology*, Vol. 1 (eds Mishra, N., Dunkel, V. & Mehlman, M.) 467–501 (Senate, New Jersey, 1980).
11. Finter, N. *J. gen Virol.* **5**, 419–427 (1969).
12. Nilsen, T. W., Wood, D. L. & Baglioni, C. *Nature* **286**, 178–180 (1980).
13. Cearniecki, C. W., Sreevalsan, T., Friedman, R. & Panet, A. *J. Virol.* **37**, 827–831 (1981).
14. Ankel, H., Krishnamurti, C., Besancon, F., Stefanos, S. & Falcoff, E. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2528–2532 (1980).
15. Stewart, W. E. II, de Clerq, E., Billiau, A., Desmeyter, J. & de Somer, P. *Proc. natn. Acad. Sci. U.S.A.* **69**, 1851–1854 (1972).
16. Creasey, A. A., Bartholomew, J. C. & Merigan, T. C. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1471–1475 (1980).
17. Gresser, I., Bandu, M. -T. & Brouty-Boyé, D. *J. natn. Cancer Inst.* **52**, 553–559 (1974).
18. Kuwata, T., Fuse, A. & Morinaga, N. *J. gen. Virol.* **33**, 7–15 (1976).
19. Lin, S. L., Ts'o, P. O. P. & Hollenberg, M. D. *Biochem. biophys. Res. Commun.* **96**, 168–174 (1980).
20. Lin, S. L. thesis, Johns Hopkins Univ. (1982).
21. Laug, W. E., Jones, P. A. & Benedict, W. F. *J. natn. Cancer Inst.* **54**, 173–179 (1975).
22. Morry, D. W. & Ts'o, P. O. P. *Mutat. Res.* **70**, 221–229 (1980).
23. Thompson, L. H. & Baker, R. M. *Meth. Cell Biol.* **6**, 209–281 (1973).
24. Davey, M. W., Sulkowski, E. & Carter, W. A. *J. biol. Chem.* **251**, 7620–7625 (1976).

Stimulation of adrenal mitogenesis by N-terminal proopiocortin peptides

F. E. Estivariz & F. Iturriza

Centro de Estudios Endocrinos, Facultad de Ciencias Medicas, Universidad Nacional de la Plata, Casilla de Correo 455, 1900 La Plata, Argentina

C. McLean, J. Hope & P. J. Lowry

Pituitary Hormone Laboratory, Department of Chemical Pathology, St Bartholomew's Hospital, London EC1 7BE UK

Although it was generally believed that corticotropin (ACTH) maintained both the size of the adrenal gland and its level of steroid production[1], there is a growing body of evidence for a specific factor(s) distinct from ACTH which may be responsible for the stimulation of adrenocortical growth and proliferation. While direct neural influences have been proposed to be responsible for the compensatory adrenal growth following unilateral adrenalectomy[2], several observations have suggested the involvement of peptides such as angiotensin[3] and vasopressin[4], which have other hormonal effects, or, for example, fibroblast growth factor[5], which has more general mitogenic actions. Moreover ACTH inhibits cell proliferation *in vitro*[6–8], and physiological doses *in vivo* cannot induce the compensatory growth and hyperplasia seen in the remaining adrenal gland after unilateral adrenalectomy[9]. These observations coupled with the lack of effect of chronic treatment with an ACTH antiserum on adrenal size[10] and the existence of adrenal weight-maintaining activity in the plasma of patients with ACTH-secreting adenomas[11], have led to the search for a pituitary factor related to ACTH which is capable of stimulating adrenal growth. We now report that peptides derived from the N-terminal non-γ-melanocyte-stimulating hormone (MSH) portion of the precursor of ACTH, proopiocortin, are potent stimulators of adrenal DNA synthesis *in vitro* and mitosis *in vivo* and may thus be involved in the physiological control of adrenal growth.

0028–0836/82/220419—04$01.00

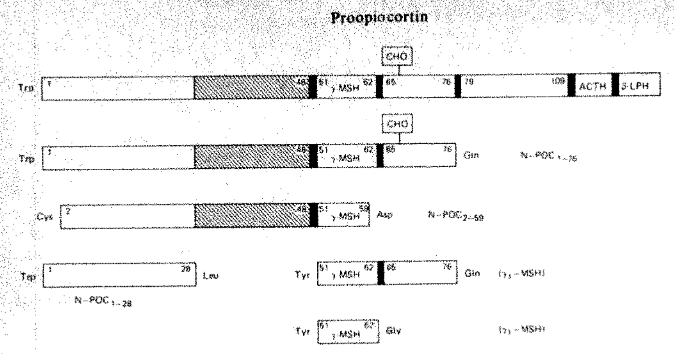


Fig. 1 Diagram of the human N-POC peptides used in this study in relation to each other and to their precursor proopiocortin. The numbering has been assigned from chemical characterization of isolated peptides[18] and cDNA[19] sequence data. The γ_2- and γ_3-MSH synthetic peptides were a gift from Dr Nicholas Ling. The hatched area corresponds to the region recognized by the antiserum[17] used in this study and the boxed CHO, the polysaccharide moiety.

ACTH is derived from a precursor which contains lipotropin (LPH)/endorphin as the C-terminal portion and an N-terminal glycopeptide (N-POC) of some 100 amino acids (Fig. 1) which contains a third melanotropin sequence (γ-MSH)[12]. A human pituitary glycopeptide (N-POC$_{1-76}$) comprising almost all of the non-ACTH non-LPH/endorphin region, which is inactive in maintaining adrenal weight and possesses weak melanotropic activity[13], has recently been found to potentiate the ACTH-induced secretion of glucocorticoids and mineralocorticoids from the adrenal cortex[14] by increasing RNA synthesis[15]. The potentiation of steroidogenesis can also be demonstrated with small synthetic peptides containing the γ-MSH sequence found in the C-terminal part of the glycopeptide[16]. Initial attempts to demonstrate adrenal weight maintenance with human N-POC$_{1-76}$ given either alone or with physiological doses of ACTH to hypophysectomized rats were unsuccessful[13].

The generation of an antiserum directed against the middle portion of this peptide (residues 29–48)[17] allowed us to test its effects on the adrenal in intact rats. Rats were injected daily for 7 days with the anti-N-POC antiserum, an antiserum which cross-reacts fully with both ACTH and β-LPH, or normal rabbit serum. Thymidine incorporation was also measured in these experiments as a more specific index of DNA synthesis. A decrease in adrenal weight (14%) was observed in the anti-N-POC antiserum-treated group although the result was not significant (results not shown). However, as shown in Fig. 2*a*, there was a significant decrease in thymidine incorporation in the anti-N-POC antiserum-treated group (B), whereas the anti-ACTH/LPH antiserum-treated group (C) did not differ significantly from controls (A). The efficacy of the anti-ACTH/LPH antiserum treatment was shown by the significantly lower levels of plasma corticosterone in this group.

In preliminary experiments, however, we had also failed to observe any effects of several days' treatment with N-POC$_{1-76}$ in rats *in vivo* using thymidine incorporation as the measure of DNA synthesis. This led us to suspect that N-POC$_{1-76}$ could be a precursor for the mitogenic factor—this indeed appeared to be the case. Figure 2*b* confirms the absence of an effect after seven daily injections of the intact N-POC$_{1-76}$ (E) but pretreatment of the peptide with trypsin revealed significant activity (F). In this experiment, dexamethazone blockade was used in order to preserve as fully as possible the endocrine status of the animals whilst avoiding the stress-induced release of ACTH-related peptides due to the daily injections. Recently[18] we have purified and chemically characterized smaller fragments from the N-terminus of N-POC$_{1-76}$, including N-POC$_{1-28}$ and N-POC$_{2-59}$, and have compared the effects of these peptides with synthetic γ_2-MSH, γ_3-MSH and ACTH on isolated adrenal cells. Whereas γ_2 and γ_3-MSH had no effect in this system

Table 1 Effect of POC peptides on adrenal growth and mitosis *in vivo*

Group	No. rats	Adrenal weight	Mitotic ratio
Control	5	28.08 ± 1.07	0.57 ± 0.21
N-POC$_{1-28}$	5	32.5 ± 0.73†	3.39 ± 0.4§
N-POC$_{1-76}$	4	28.85 ± 0.87 NS	0.81 ± 0.32
ACTH$_{1-24}$	5	30.86 ± 1.01*	2.12 ± 0.35‡

Adrenal weight was calculated per 100 g body weight and mitotic ratio was the number of mitoses per mm^2 of adrenal cortex (containing ~4,000 cells)[28]. All groups were compared with control values. The difference in mitotic ratio between N-POC$_{1-28}$- and ACTH$_{1-24}$-treated animals was significant ($P<0.05$). There was no significant variation between the groups with respect to whole body weight. Ten week old female Sprague–Dawley rats were implanted subcutaneously with Alzet osmotic minipumps (model 2001) containing 1 mM acetic acid (control) or 1 mM acetic acid containing 100 μg ml^{-1} of the appropriate peptide and delivered to each rat at ~3 μg of peptide per day. Rats were distributed randomly between cages and were allowed food and water *ad libitum*. On day 7, each rat was injected with colchicine (1 μg per g body weight) and killed 6 h later. Adrenals were removed, weighed and fixed immediately in Bouins, prepared for histology in the usual way and stained using Feulgen's method. Three mid-sections of whole cortex of each adrenal (that is, a total of 30 sections per group) were examined microscopically to calculate mitotic ratios. NS, not significant.
* $P<0.05$; † $P<0.02$; ‡ $P<0.01$; § $P<0.001$.

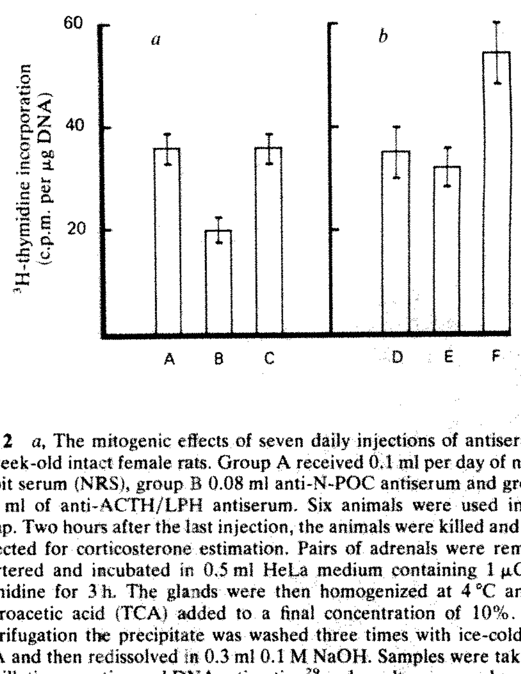


Fig. 2 *a*, The mitogenic effects of seven daily injections of antisera into 3-week-old intact female rats. Group A received 0.1 ml per day of normal rabbit serum (NRS), group B 0.08 ml anti-N-POC antiserum and group C 0.05 ml of anti-ACTH/LPH antiserum. Six animals were used in each group. Two hours after the last injection, the animals were killed and blood collected for corticosterone estimation. Pairs of adrenals were removed, quartered and incubated in 0.5 ml HeLa medium containing 1 μCi ^{3}H-thymidine for 3 h. The glands were then homogenized at 4 °C and trichloroacetic acid (TCA) added to a final concentration of 10%. After centrifugation the precipitate was washed three times with ice-cold 10% TCA and then redissolved in 0.3 ml 0.1 M NaOH. Samples were taken for scintillation counting and DNA estimation[29] and results expressed as c.p.m. per μg DNA. Group B is significantly different from both the NRS and anti-ACTH/LPH antiserum groups ($P<0.01$). Corticosterone concentrations (μg per 100 ml) in each group were: A, 15.2 ± 3.1; B, 13.4 ± 1.9; C, 9.4 ± 12. The anti-ACTH/LPH antiserum was raised in sheep by immunization with ACTH$_{1-24}$ coupled to bovine thyroglobulin by glutaraldehyde. Using radioiodinated ACTH$_{1-39}$ as label, α-MSH, β-MSH, β-LPH and γ-LPH showed full molar cross-reactivity in a radioimmunoassay using this antiserum at a final dilution of 1:12,000. *b*, The mitogenic effects of N-POC$_{1-76}$ and trypsinized N-POC$_{1-76}$ on the adrenals of 7-week-old female rats. Ten animals were used in each group and dexamethasone blockade was achieved by adding the compound (10 μg ml^{-1}) to the drinking water at the start of the experiment. Group E received seven daily injections of 5 μg N-POC in 10 μl saline. In group F, N-POC$_{1-76}$ (0.5 mg) was treated with DCC-trypsin at a concentration of 1:200 (enzyme/peptide) for 2 h in 1 ml 10 mM NaHCO$_3$ at 37 °C, and was then diluted 10 times with saline and stored at −20 °C. Rats received the same amount of peptide as in group E. The saline used for the control group (D) had incubated, heat-treated DCC-trypsin added to it equivalent to that used to digest the injection material in group F. Two hours after the last injection, the animals were killed and the adrenals treated as in *a*. Thymidine incorporation in group F was significantly higher than in control (D) and intact N-POC$_{1-76}$-treated groups ($P<0.05$). Corticosterone concentrations (μg per 100 ml) were significantly depressed in all groups: D, 6.8 ± 1.2; E, 7.4 ± 0.6; F, 8.1 ± 0.8.

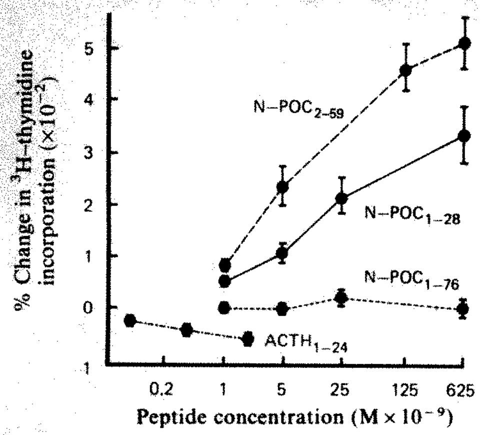


Fig. 3 Dose-response curves of N-POC peptide-mediated stimulation of ^{3}H-thymidine incorporation into DNA of dispersed rat adrenal cells. Cells were prepared as described previously[30]. Approximately 10^5 cells were distributed into each vial containing 0.5 ml of HeLa medium containing 0.2% bovine serum albumin (BSA) and 0.2% lima bean trypsin inhibitor (LBI). Cells were preincubated for 3 h before addition of N-POC peptides. After a further 3 h incubation, 1 μCi of ^{3}H-thymidine was added to each vial and incorporation was allowed to proceed for 6 h. Cells were then centrifuged (100*g*, 10 min), washed with fresh medium and sonicated in 0.5 ml ice-cold 12% TCA after addition of 100 μg BSA to each vial as carrier. After washing with 10% TCA the precipitate was redissolved in 0.1 M NaOH and treated as for Fig. 2. Control cells gave 2,108 ± 235 c.p.m.

(results not shown), both N-POC$_{1-28}$ and N-POC$_{2-59}$, at 10^{-9}–10^{-7} M, elicited a dose-dependent incorporation of thymidine into DNA, the larger peptide being more potent (Fig. 3). N-POC$_{1-76}$ had little effect at the highest dose assayed and ACTH inhibited thymidine incorporation in a dose-related manner. The synthetic γ-MSHs are very weak steroidogenic agents and N-POC$_{1-28}$ and N–POC$_{2-59}$ are devoid of this activity in the isolated adrenal cell bioassay.

The availability of the smaller N-POC peptides also allowed us to test their adrenal action *in vivo* and to exclude the possibility of artefacts in the thymidine incorporation assay caused by possible changes in the endogenous adrenal thymidine pool size in treated animals. In these experiments we used implanted minipumps to administer the peptides and thus achieved a continuous stimulus over 7 days while completely preserving the endocrine status of the animal. After 1 week's treatment there was no significant difference between the N-POC$_{1-76}$ and control groups with regard to either adrenal weight or mitotic index (Table 1). The N-POC$_{1-28}$-treated group, however, showed a highly significant increase in adrenal weight and mitotic index and seemed more potent than the ACTH$_{1-24}$ group in both respects (Table 1).

Thus although ACTH can be shown to be mitogenic *in vivo* (ref. 1 and Table 1), it initiates a different process of cellular growth *in vivo* and inhibits the hyperplasia seen in the remaining adrenal after unilateral adrenalectomy[9]. It also inhibits cell proliferation *in vitro* (refs 6–8 and Fig. 3). This, coupled with the observations that anti-ACTH antiserum treatment does not lead to adrenal atrophy (ref. 10 and Fig. 2*a*) or affect the compensatory adrenal hyperplasia in unilateral adrenalectomy[10], argue against ACTH as the physiological mitogen. However, the peptides derived from the N-terminal non-γ-MSH region of proopiocortin used in the present study were active both *in vivo* and *in vitro* and thus may resemble the major adrenal mitogenic hormones. The extensive conservation previously noted in the N-POC region from cDNA sequence data of rat[19], human[20] and bovine material is confirmed in a recent comparative immunohistochemical survey, which shows that the anti-N-POC antiserum used in our study immunostains only the corticotropes and melanotropes of a variety of vertebrate species including fish[21]. This also suggests a biological role for this region.

The need for cleavage of the N-POC region before its mitogenic activity can be expressed suggests an interesting control mechanism. Chemical[13], biosynthetic and physiological studies[22] have indicated that the main molecular form of the corticotropin in blood may be similar to the intact N-POC$_{1-76}$, which is inactive as an adrenal mitogen. The antiserum experiments, however, would suggest either secretion of smaller fragments of N-POC, for example from the pars intermedia (where limited further processing of this region has been shown to take place), or the release of a peptidase which cleaves the peptide after secretion. The most probable cleavage site would be at the double basic sequence at residues 49–50 giving rise to N-POC$_{1-48}$ and N-POC$_{50-76}$ (γ_3-MSH in the ox[23] and rat[24] has been found to start with the lysine residue at position 50). Thus the most likely physiological fragment would be N-POC$_{1-48}$. This is corroborated by the observation that the antiserum used in this study effectively removed the naturally circulating biological activity and did not cross-react with smaller biologically active fragments (N-POC$_{1-28}$) in the radioimmunoassay.

In mammals with little or no pars intermedia tissue or its associated peptides, such as humans[25,26], post-secretional modification would be the probable mechanism. Thus a change in processing of the N-POC$_{1-76}$ region could provide the necessary mitogenic stimulus to cause, for example, the onset of the adrenarche and development of the androgenic zone of the prepubertal human adrenal gland. As the secretion of peptides derived from N-POC in the corticotrope has been shown to be under the same hypothalamic control as ACTH and β-LPH/endorphin[27], their depressed levels during chronic treatment with glucocorticoids could be responsible for the adrenal atrophy observed in such cases. Administration of physiological or even pharmacological doses of peptides derived from this region, whilst continuing glucocorticoid therapy, could provide the necessary mitogenic stimulus and lead to a rapid recovery of the functioning of the hypothalamic pituitary adrenal axis on cessation of that therapy.

Thus we have the unique situation where the N-terminal region of proopiocortin, including ACTH, provides a variety of peptides capable of affecting different aspects of the adrenal gland's activity. N-POC$_{1-28}$, N-POC$_{2-59}$ and ACTH stimulate DNA synthesis, the intact N-POC$_{1-76}$ potentiates the action of ACTH by causing RNA synthesis[15], N-POC$_{51-76}$ activates cholesterol ester hydrolase[16] and ACTH switches on the translation of both the stable[1] and N-POC$_{1-76}$-stimulated mRNA pools[10] required for the initiation of corticosteroidogenesis.

The technical assistance of Mrs M. Carino is acknowledged. The work was supported by the Commision de Investigaciones Cientificas de la Provincia de Buenos Aires and by the MRC.

Received 28 October 1981; accepted 2 April 1982.

1. Garren, L. D., Gill, G. N., Masui, H. & Walton, G. M. *Recent Prog. Horm. Res.* **27**, 433–474 (1971).
2. Dallman, M. F., Engeland, W. C. & Shinsako, J. *Am. J. Physiol.* **231**, 408–414 (1976).
3. Gill, G. N., Ill, C. R. & Simonian, M. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5569–5573 (1977).
4. Payet, N. & Lehoux, J. G. *J. Steroid Biochem.* **12**, 461–467 (1980).
5. Gospodarowicz, D. & Handley, H. H. *Endocrinology* **97**, 102–107 (1975).
6. O'Hare, M. J. & Munro Neville, A. *Biochem. Soc. Trans.* **1**, 1088–1091 (1973).
7. Masui, H. & Garren, L. D. *Proc. natn. Acad. Sci. U.S.A.* **68**, 3206–3210 (1971).
8. Ramachandran, J. & Suyama, A. T. *Proc. natn. Acad. Sci. U.S.A.* **72**, 113–117 (1976).
9. Dallman, M. F., Engeland, W. C., Holzwarth, M. A. & Scholz, P. M. *Endocrinology* **107**, 1399–1404 (1980).
10. Rao, A. J., Long, J. A. & Ramachandran, J. *Endocrinology* **102**, 371–387 (1978).
11. Segal, B. M. & Christy, N. P. *J. clin. Endocr.* **28**, 1465–1472 (1968).
12. Nakanishi, S. *et al. Nature* **278**, 423–427 (1979).
13. Estivariz, F. E., Hope, J., McLean, C. & Lowry, P. J. *Biochem. J.* **191**, 125–132 (1980).
14. Al-Dujaili, E. A. S., Hope, J., Estivariz, F. E., Lowry, P. J. & Edwards, C. R. W. *Nature* **291**, 156–159 (1981).
15. Al-Dujaili, E. A. S., Williams, B. C., Edwards, C. R. W., Salacinski, P. R. & Lowry, P. J. *Biochem. J.* **204**, 301–305 (1982).
16. Pedersen, R. C., Brownie, A. C. & Ling, N. *Science* **208**, 1044–1045 (1980).
17. Hope, J., Ratter, S. J., Estivariz, F. E., McLoughlin, L. & Lowry, P. J. *Clin. Endocr.* **15**, 221–227 (1981).
18. McLean, C., Hope, J., Salacinski, P. R., Estivariz, F. E. & Lowry, P. J. *Biosci. Rep.* **1**, 843–849 (1981).
19. Drouin, J. & Goodman, H. M. *Nature* **288**, 610–618 (1980).
20. Chang, A. C. Y., Cochet, M. & Cohen, S. N. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4890–4894 (1980).
21. Estivariz, F. E., Iturriza, F. C., Hope, J. & Lowry, P. J. *Gen. comp. Endocr.* **46**, 1–6 (1982).

22. Eipper, B. A. & Mains, R. E. *J. supramolec. Struct.* **8**, 247–262 (1978).
23. Ling, N., Shibaski, T., Esch, F. & Bohlen, P. *63rd A. Meet. Am. Endocrine Soc.* Abstr. no. 383 (The Endocrine Society, 1981).
24. Browne, C. A., Bennett, H. P. J. & Solomon, S. *Biochemistry* **20**, 4538–4546 (1981).
25. Wingstrand, K. G. *The Pituitary Gland* Vol. 3 (eds Harris, G. W. & Donovan, B. T.) 1–27 (Butterworths, London, 1966).
26. Silman, R. E. *et al. Nature* **260**, 716–718 (1976).
27. Estivariz, F. E., Gillies, G. E. & Lowry, P. J. *Pharmac. Ther.* **13**, 61–67 (1981).
28. Van Dorp, A. W. P. & Dean, H. W. *Anat. Rec.* **107**, 265–282 (1950).
29. Giles, K. W. & Myers, A. *Nature* **206**, 93 (1965).
30. Lowry, P. J., McMartin, C. & Peters, J. *J. Endocr.* **59**, 43–55 (1973).

N-methyl-D-aspartate-type receptors mediate striatal ^{3}H-acetylcholine release evoked by excitatory amino acids

Bernard Scatton & John Lehmann*

Synthelabo-LERS, 31 Avenue Paul Vaillant Couturier, 92220 Bagneux, France

Several lines of evidence suggest that striatal cholinergic interneurones receive an excitatory input from the cerebral cortex which utilizes an excitatory amino acid, L-glutamate or L-aspartate, as its neurotransmitter. Cortical ablation reduces striatal high-affinity glutamate uptake[1,2] and concomitantly decreases acetylcholine turnover[3]. The destruction of cholinergic, as well as other, neurones of the striatum by the rigid glutamate analogue kainic acid seems to require a functional excitatory amino acid innervation[4,5]. Furthermore, electron microscopic studies[6] suggest the existence of a cortical input to the aspiny dendrites of neurones morphologically similar to the presumed striatal cholinergic interneurones[7,8]. The recent discovery of preferential antagonists of amino acid-induced excitation in the vertebrate central nervous system has led to the classification of excitatory amino acid receptors into three subtypes: the *N*-methyl-D-aspartate (NMDA), the quisqualate- and the kainate-preferring receptors[9]. We have now attempted to characterize the receptor(s) mediating the excitatory amino acid influence on striatal cholinergic neurones by investigating the effects of specific excitatory amino acid receptor agonists and antagonists on the release of ^{3}H-acetylcholine from slices of the rat corpus striatum. We find that excitatory amino acid analogues evoke a tetrodotoxin-sensitive release of ^{3}H-acetylcholine from rat striatal slices superfused in Mg^{2+}-free medium, NMDA and ibotenate being the most potent and kainate and quisqualate the least potent. This effect is antagonized by Mg^{2+} and by (−)2-amino-7-phosphonoheptanoate (−APHept) and (±)2-amino-5-phosphonopentanoate (±APPent) but not by glutamic acid diethyl ester (GDEE). These data suggest the involvement of a NMDA-type receptor in the excitatory amino acid influence on striatal acetylcholine release.

The influence of excitatory amino acid receptor agonists on the release of striatal ^{3}H-acetylcholine was first investigated in normal Krebs medium (containing 1.2 mM Mg^{2+}). Exposure to 100 μM L-glutamate, L-aspartate and the excitatory amino acid receptor agonists, *N*-methyl-DL-aspartate, (±)ibotenate, *RS*-α-amino-3-hydroxy-5-methyl-4-isoazole-propionic acid (AMPA), L-cysteate, kainate and quisqualate of superfused rat striatal slices previously incubated with ^{3}H-choline evoked a moderate release of ^{3}H-acetylcholine (Table 1). AMPA and kainate were the most effective whereas *N*-methyl-DL-aspartate was almost ineffective. In the absence of calcium, L-glutamate-evoked ^{3}H-acetylcholine release was abolished (data not shown).

As electrophysiological studies on frog and rat spinal neurones have indicated that the different types of excitatory amino acid receptors exhibit a differential sensitivity to Mg^{2+} (low concentrations of Mg^{2+} antagonize NMDA- or L-aspar-

* Present address: Department of Neuroscience, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA.

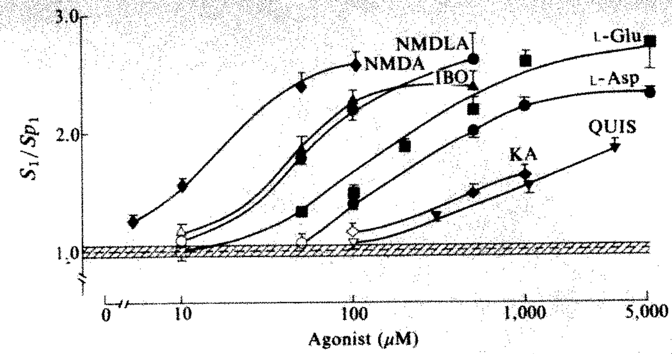


Fig. 1 Effects of excitatory amino acid receptor agonists on the outflow of ^{3}H-acetylcholine from striatal slices in Mg^{2+}-free medium: concentration dependence. Methods were as described in Table 1 legend. Each value represents the mean with s.e.m. of eight independent determinations (performed on eight rats). The dotted line and shaded area indicate the basal level with s.e.m. of the release of ^{3}H-acetylcholine in controls ($S_1/Sp_1 = 0.99 \pm 0.05$). Filled symbols: statistically significant ($P < 0.001$) compared with controls (Student's *t*-test). Open symbols: not statistically significant. IBO, (±)ibotenate; NMDA, *N*-methyl-D-aspartate; NMDLA, *N*-methyl-DL-aspartate; L-GLU, L-glutamate; L-ASP, L-aspartate; KA, kainate; QUIS, quisqualate.

Table 1 Effects of excitatory amino acid receptor agonists on outflow of ^{3}H-acetylcholine from striatal slices in medium containing 1.2 mM Mg^{2+} and in Mg^{2+}-free medium

	S_1/Sp_1	
Agonist (100 μM)	With Mg^{2+}	Without Mg^{2+}
Controls	0.93 ± 0.01	0.94 ± 0.02
N-methyl-DL-aspartate	0.97 ± 0.03	2.12 ± 0.03†‡
(±)Ibotenate	1.05 ± 0.05*	2.07 ± 0.05†‡
AMPA	1.16 ± 0.03†	1.63 ± 0.03†‡
L-glutamate	1.11 ± 0.03†	1.63 ± 0.05†‡
L-aspartate	1.09 ± 0.02†	1.43 ± 0.05†‡
L-cysteate	1.02 ± 0.03*	1.33 ± 0.04†‡
Kainate	1.14 ± 0.03†	1.19 ± 0.02†
Quisqualate	1.03 ± 0.03†	1.06 ± 0.03†

Experiments were performed on male Sprague–Dawley rats (COBS CD Charles River, France, 150 g). Striatal slices (1×1×0.3 mm prepared by using a McIlwain tissue chopper) were incubated at 37 °C with 0.05 μM [*N*-methyl-^{3}H]choline (specific activity 80 Ci mmol^{-1}; NEN) in oxygenated Krebs medium (composition in mM: NaCl 118, KCl 4.2, $NaHCO_3$ 25, NaH_2PO_4 1.0, $CaCl_2$ 2.6, glucose 11, $MgCl_2$ 1.2) for 30 min to allow high-affinity uptake to occur. The slices were subsequently washed, transferred to a Plexiglass chamber (8 mm diameter × 6 mm) and continuously superfused at a constant rate (0.5 ml min^{-1}) with normal or Mg^{2+}-free Krebs medium containing 10 μM hemicholinium-3 (a choline uptake inhibitor). The superfusate was collected every 6 min. Excitatory amino acid receptor agonists (100 μM) were added to the superfusion medium for 2 min at 48 min after the beginning of the superfusion. At the end of the experiment, slices were recovered and dissolved in 500 μl 0.5 M sodium hydroxide for radioactivity measurement. Experiments performed in the presence of eserine (30 μM) revealed that the radioactivity released in the medium consisted of 92% authentic ^{3}H-acetylcholine. For each fraction the amount of radioactivity released into the medium as a percentage of total tissue content during that interval (fractional release) was calculated. The elevation of ^{3}H-acetylcholine outflow caused by the agonists is expressed as the ratio of the fractional release obtained in the 6 min interval after the beginning of the 2 min pulse of agonist (S_1) divided by the fractional release obtained in the 6 min period preceding the 2 min pulse of agonist (Sp_1). Each value is the mean with s.e.m. of eight independent determinations (performed on eight animals). The statistical significance of the data was evaluated by the Student's two-tailed *t*-test.

* $P < 0.05$; † $P < 0.01$ compared with controls; ‡ $P < 0.001$ compared with Mg^{2+}-containing medium.

0028–0836/82/220422—03$01.00

tate-induced but not kainate- and quisqualate-induced depolarization)[9], the responses evoked by excitatory amino acid receptor agonists (100 μM) were studied in Mg^{2+}-free conditions. When Mg^{2+} was removed from the medium, the ability of *N*-methyl-DL-aspartate to release ^{3}H-acetylcholine increased dramatically (Table 1). Removal of Mg^{2+} also potentiated the response elicited by (±)ibotenate, AMPA, L-glutamate, L-aspartate and L-cysteate but not that elicited by kainate. This effect was greatest for the agonists which were more specific for NMDA-type receptors (*N*-methyl-DL-aspartate, ibotenate). The relative potencies of excitatory amino acid agonists on the release of ^{3}H-acetylcholine in Mg^{2+}-free medium were compared by plotting concentration–response curves. All the curves were sigmoidal in shape, suggesting the involvement of a single saturable receptor (Fig. 1). The relative potency of the agonists was the following: NMDA > (±)ibotenate ≥ *N*-methyl-DL-aspartate > *L*-glutamate > *L*-aspartate > kainate ≥ quisqualate. This suggests that the receptor which mediates the release of ^{3}H-acetylcholine evoked by excitatory amino acid agonists shows a preference for NMDA-type agonists.

As suggested by electrophysiological studies, excitatory amino acid receptor subtypes exhibit a differential sensitivity to blockage by various organic antagonists. The most potent and selective NMDA-type receptor antagonists are the phosphonate analogues of carboxylic acids (−)APHept and (±)APPent, whereas GDEE is considered an antagonist of quisqualate-type receptors[10,11]. We further characterized the receptor involved in excitatory amino acid-evoked release of ^{3}H-acetylcholine by testing the effects of the amino acid antagonists on the response evoked by *N*-methyl-DL-aspartate and

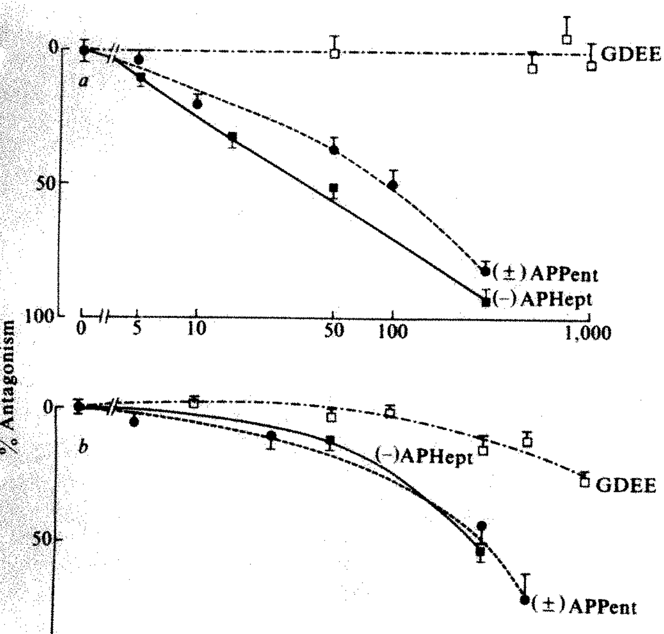


Fig. 2 Effects of excitatory amino acid receptor antagonists on *a*, *N*-methyl-DL-aspartate- and *b*, L-glutamate-evoked release of ^{3}H-acetylcholine from rat striatal slices. Methods were as described in Table 1 legend. The amino acid antagonist in the concentration indicated was added to the superfusion medium 12 min before the 2 min pulse (S_1) of *N*-methyl-DL-aspartate (50 μM) or L-glutamate (300 μM). The ratio S_1/Sp_1 was calculated as described in Table 1 legend in both the absence and presence of the antagonist and results were expressed as per cent of this value in respective controls (agonist alone). Results are means with s.e.m. of eight independent determinations (performed on eight animals). Filled symbols: statistically significant (at $P<0.01$) compared with respective controls (Student's *t*-test). Open symbols: not statistically significant.

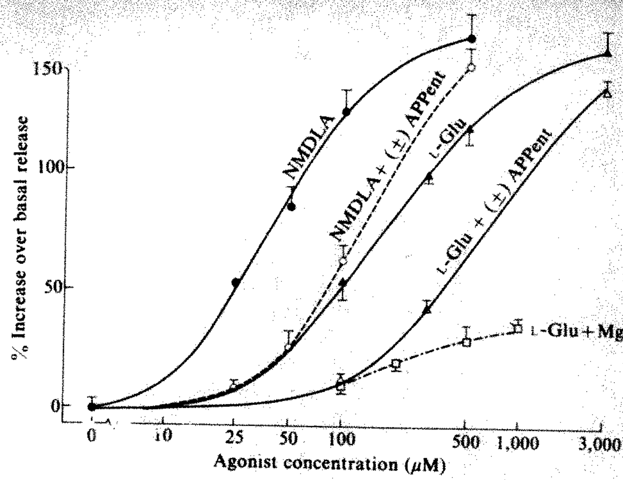


Fig. 3 Competitive antagonism by (±)APPent and non-competitive antagonism by Mg^{2+} of the N-methyl-DL-aspartate- and L-glutamate-evoked release of striatal ^{3}H-acetylcholine. For methods and expression of results see Table 1 legend. Mg^{2+} (1.2 mM) or (±)APPent (50 and 300 μM when tested against NMDLA and L-glutamate respectively) were introduced in the superfusion medium 12 min before the 2 min pulse of the excitatory amino acid agonist. Each point corresponds to the mean with s.e.m. of eight independent determinations (performed on eight animals).

L-glutamate in Mg^{2+}-free conditions. (−)APHept and (±)APPent antagonized in a concentration-dependent manner the release of ^{3}H-acetylcholine evoked by *N*-methyl-DL-aspartate (50 μM) with IC_{50}s of 40 and 90 μM respectively (Fig. 2). Both drugs also antagonized the response to L-glutamate (300 μM) with IC_{50}s of 280 and 320 μM respectively. In contrast, the quisqualate receptor antagonist GDEE failed to affect the response evoked by *N*-methyl-DL-aspartate and only weakly antagonized (as a concentration of 1 mM) the L-glutamate response. The inhibition by (±)APPent of the *N*-methyl-DL-aspartate- and L-glutamate evoked release of ^{3}H-acetylcholine is clearly competitive (Fig. 3), in contrast to the Mg^{2+} inhibition which seems to be non-competitive. In the concentration range tested, neither (−)APHept nor (±)APPent alone affected the spontaneous outflow of ^{3}H-acetylcholine. Moreover, at a concentration effective in blocking *N*-methyl-DL-aspartate or L-glutamate responses, (±)APPent only marginally inhibited the release of ^{3}H-acetylcholine evoked by other depolarizing agents—K^+ (12 mM) and veratridine (1 μM) S_1/Sp_1: controls, 0.96±0.02; K^+, 2.74±0.06; K^++(±)APPent, 2.30±0.01, $P<0.01$, Student's two-tailed *t*-test; veratridine, 1.70±0.09; veratridine+(±)APPent, 1.61±0.04, $n=8$), suggesting that the antagonism occurs at the level of specific amino acid receptors.

The present data are consistent with the involvement of an NMDA-type receptor in the excitatory amino acid influence on striatal ^{3}H-acetylcholine release. This conclusion is supported by (1) the high potency of the NMDA-type receptor agonists NMDA and (±)ibotenate to evoke the release of the labelled transmitter in Mg^{2+}-free conditions; (2) the antagonism by low concentrations of Mg^{2+} of NMDA-type agonist-evoked release of striatal ^{3}H-acetylcholine; (3) the selective blockage by antagonists of NMDA-type receptors of the *N*-methyl-DL-aspartate- and L-glutamate-evoked release of ^{3}H-acetylcholine in Mg^{2+}-free medium. Note that NMDA has also been found to be the most potent excitatory amino acid in stimulating sodium efflux from striatal slices[12]. The very weak potency of quisqualate, together with the failure of the quisqualate-type antagonist GDEE to block *N*-methyl-DL-aspartate- or L-glutamate-evoked ^{3}H-acetylcholine release in both Mg^{2+}-free conditions and Mg^{2+}-containing medium (not shown), suggest that quisqualate-type receptors do not contribute significantly to the excitatory amino acid-evoked responses of striatal cholinergic cells.

The localization of the NMDA receptors mediating ^{3}H-acetylcholine release evoked by excitatory amino acid agonists is not yet known. However, the observation that addition of tetrodotoxin (0.5 μM) abolished the *N*-methyl-DL-aspartate (50 μM) and L-glutamate (300 μM)-evoked release of ^{3}H-acetylcholine in Mg^{2+}-free medium (S_1/Sp_1: controls, 0.96±0.02; *N*-methyl-DL-aspartate, 2.88±0.06; *N*-methyl-DL-aspartate+tetrodotoxin, 1.10±0.07*; L-glutamate, 2.75±0.02; L-glutamate+tetrodotoxin, 1.09±0.03*; $*P<0.01$, Student's *t*-test, compared with agonist alone) indicates that action potentials mediate the release of ^{3}H-acetylcholine evoked by these agonists. These data are consistent with the hypothesis that excitatory amino acid receptor agonists activate receptors located on cholinergic dendrites, causing depolarization and action potential propagation along the axons of the cholinergic interneurones to the terminals, where ^{3}H-acetylcholine is released. The involvement of excitatory amino acid receptors on cells other than cholinergic neurones which affect the cholinergic neurones via an unknown transmitter cannot, however, be excluded rigorously.

In conclusion, an NMDA-type receptor has been found to mediate the excitatory influence of amino acids on striatal cholinergic neurones. These receptors, which may be located on dendrites of cholinergic cells, may play a part in transducing information received from the cerebral cortical afferents, which are thought to use an excitatory amino acid as neurotransmitter.

We thank Dr J. F. Collins, P. Krogsgaard-Larsen and J. C. Watkins for gifts of compounds, and D. Fage for technical assistance. (−)APHept was obtained from Dr Collins, Department of Chemistry, City of London Polytechnic, and (±)APPent from Dr Watkins, Department of Pharmacology and Physiology, The Medical School, Bristol.

Received 29 January; accepted 14 April 1982.

1. McGeer, P. L., McGeer, E. G., Scherer, U. & Singh, K. *Brain Res.* **128,** 369–373 (1977).
2. Divac, I., Fonnum, F. & Storm-Mathisen, J. *Nature* **266,** 377–378 (1977).
3. Wood, P. L., Moroni, F., Cheney, D. L. & Costa, E. *Neurosci. Lett.* **12,** 349–354 (1979).
4. McGeer, E. G., McGeer, P. L. & Singh, K. *Brain Res.* **139,** 381–383 (1978).
5. Biziere, K. & Coyle, J. T. *Neurosci. Lett.* **8,** 303–310 (1978).
6. Hassler, R., Chung, J. W., Rinne, U. & Wagner, A. *Expl Brain Res.* **31,** 67–80 (1978).
7. Lehmann, J. & Fibiger, H. C. *Life Sci.* **25,** 1939–1947 (1979).
8. Kimura, H., McGeer, P. L., Peng, F. & McGeer, E. G. *Science* **208,** 1057–1059 (1980).
9. Watkins, J. C. in *Glutamate: Transmitter in the Central Nervous System* (eds Roberts, P. J., Storm-Mathisen, J. & Johnston, G. A. R.) 1–24 (Wiley, Chichester, 1981).
10. Watkins, J. C., Davies, J., Evans, R. H., Francis, A. A. & Jones, A. W. in *Glutamate as a Neurotransmitter* (eds Di Chiara, G. & Gessa, G. L.) 263–273 (Raven, New York, 1981).
11. Perkins, M. N., Stone, T. W., Collins, J. F. & Curry, K. *Neurosci. Lett.* **23,** 333–336 (1981).
12. Luini, A., Goldberg, O. & Teichberg, V. I. *Proc. natn. Acad. Sci. U.S.A.* **78,** 3250–3254 (1981).

Modulation of lateral mobility of band 3 in the red cell membrane by oxidative cross-linking of spectrin

Deborah K. Smith & Jiri Palek

The Department of Research, St Elizabeth's Hospital, Tufts University School of Medicine, Boston, Massachusetts 02135, USA

Several recent studies have shown that the lateral diffusion of erythrocyte transmembrane proteins is modulated by the submembrane network of spectrin, actin and polypeptide 4.1 (refs 1–3). The rate of lateral diffusion of the major transmembrane glycoprotein (the anion transport channel, band 3) is increased approximately 50-fold in membranes from spherocytic erythrocytes of mice which are deficient in the membrane skeletal protein spectrin[1]. Lateral mobility is increased when spectrin is removed from the membrane by extraction at low ionic strength[2] or when it is displaced by a fragment of ankyrin containing the membrane attachment site for spectrin[3]. Such experiments show that the presence of spectrin on the membrane restricts the mobility of membrane proteins, but they do not define the mechanism by which spectrin could modulate membrane protein mobility *in situ*. The oxidant diamide has been shown to cross-link spectrin via intermolecular disulphide coupling into a high molecular weight complex[4] and this is associated with inhibition of discocyte–echinocyte shape change[5] and decreased erythrocyte deformability[6]. Here we report that the lateral mobility of band 3 protein is decreased in erythrocytes in which spectrin is cross-linked via diamide. Furthermore, this inhibition of mobility is reversed when spectrin oxidative cross-links are reduced by addition of dithiothreitol (DTT).

Membranes isolated from red cells treated with 2 mM diamide or diamide followed by DTT were subjected to SDS-polyacrylamide gel electrophoresis in non-reducing conditions (Fig. 1). Membranes of cells treated with diamide showed the presence of a high molecular weight polymer ($>10^6$ molecular weight) which was reducible by addition of DTT to the gel. The polymer was absent in cells treated with DTT, indicating that it was also reversible in intact cells. Diamide-treated samples were electrophoresed in the second dimension in reducing conditions to identify components of the high molecular weight polymer. As shown in Figure 2, the polymer consisted primarily of spectrin (>90%) with small amounts of other membrane proteins. There was some spectrin dimer and some band 3 dimer, which might be expected because these proteins probably exist as tetramers in the membrane[7,8]. Cross-linked spectrin dimers were routinely detected in heavily loaded non-reducing gels, such as in the first-dimension gel of Fig. 2. There was no evidence, however, of cross-linking of band 3 to other membrane components such as ankyrin or spectrin, which might affect band 3 mobility directly. Therefore, in these conditions and in intact cells, diamide cross-linked spectrin fairly specifically. More extensive treatment with diamide (>2 mM)

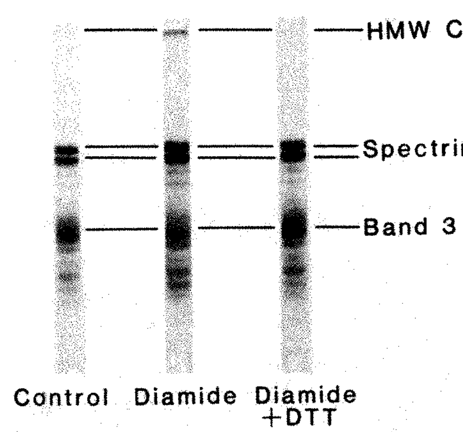


Fig. 1 Non-reducing SDS-polyacrylamide gels of membranes from diamide-treated erythrocytes. Blood was collected from healthy volunteers. Cells were washed in saline and buffy coats removed. Erythrocytes were incubated with 2 mM diamide in 5 mM Na_2HPO_4, 145 mM NaCl, *p*H 7.4 for 30 min at 37 °C. Cells were washed with the same buffer, and an aliquot of diamide-treated cells was incubated for 30 min at 37 °C with 15 mM DTT. Cells were washed, membrane ghosts prepared and samples were electrophoresed on 2.5% acrylamide–1% agarose gels as described previously[13]. All gels were stained with Coomassie brilliant blue. HMW Cx., high molecular weight complex.

caused significant cross-linking of other membrane proteins as well as spectrin (not shown).

To determine band 3 lateral diffusion rates, erythrocytes were labelled with the fluorescent molecule 5-(3,5-dichlorotriazinyl) aminofluorescein (DCTAF) in conditions where band 3 protein was specifically labelled[1]. Labelled cells were then fused with unlabelled cells by the addition of polyethylene glycol, and fused red cell doublets were observed in the fluorescence microscope for spread of fluorescence from the labelled to the unlabelled half. Label equilibrated significantly more slowly in erythrocytes in which spectrin was oxidatively cross-linked with diamide (Fig. 3). The time necessary for equilibration of label on 90% of all doublets (Fig. 3) was used to calculate a minimum diffusion constant (D_{min}) for band 3 according to the procedures of Fowler and Branton[9] and Huang[10]. D_{min} for diffusion of band 3 in normal red cell membranes was calculated to be $3\times10^{-11}\ cm^2\ s^{-1}$, which agrees with previously published values obtained by this and other methods[1-3,9,11]. D_{min} for diamide-treated cells was $\sim1.8\times10^{-11}\ cm^2\ s^{-1}$, reflecting a 40% reduction in the rate of band 3 diffusion ($P<0.005$). In diamide-treated cells which were further incubated with DTT to reduce oxidative cross-links, D_{min} was the same as for untreated cells. In experiments where DTT incompletely reduced oxidative cross-links, the extent of reversal of diamide effects on mobility were proportional to the degree of DTT reduction of the high molecular weight membrane polymer formed by diamide (not shown). These results indicate that stabilization of spectrin in a high molecular weight complex slows band 3 diffusion without cross-linking band 3 to the membrane skeleton.

Membranes isolated from red cells after fusion and equilibration of label showed no proteolysis when analysed on SDS-acrylamide gels. Furthermore, addition of the protease inhibitors phenylmethylsulphonyl fluoride (1 mM), *N*-α-*p*-tosyl-L-lysine chloromethyl ketone (1 mM) and pepstatin A ($1\ \mu g\ ml^{-1}$) had no effect on mobility of band 3 in these experiments.

Previous studies showed that the membrane skeleton serves to restrict lateral movement of transmembrane proteins, and two mechanisms have been proposed. First, studies by Golan and Veatch[2] and by Fowler and Bennett[3] on the dissociation of spectrin from the membrane suggested that the diffusion rate

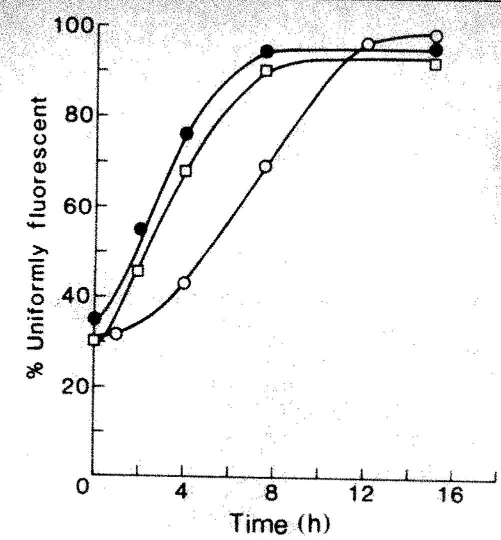


Fig. 3 Effect of diamide-induced oxidation of spectrin on lateral mobility of band 3 protein at 30 °C. Erythrocytes at 20% haematocrit in 10 mM $NaHCO_3$, 145 mM NaCl, *p*H 9.5, were incubated on ice for 1 h with $2\ mg\ ml^{-1}$ DCTAF. Cells were washed twice in 10 mM $NaHCO_3$, 95 mM NaCl, 50 mM glycine, *p*H 9.5 to remove unreacted DCTAF, then washed twice in phosphate-buffered saline (10 mM Na_2HPO_4, 145 mM NaCl, *p*H 7.4). Unlabelled cells were incubated with the same buffers in the absence of DCTAF. The labelling procedure caused no haemolysis. Labelled and unlabelled cells were fused with polyethylene glycol as described earlier[11]. Fused doublets were observed at 30 °C for diffusion and equilibration of label over the entire doublet. Data are plotted as per cent of labelled doublets which are uniformly fluorescent versus time[9]. ○, Cells treated with 2 mM diamide at 37 °C for 30 min before labelling and fusion. ●, Cells treated with diamide, washed, then incubated with 15 mM DTT at 37 °C for 30 min before labelling and fusion. □, Cells incubated at 37 °C with no diamide or DTT before labelling or fusion.

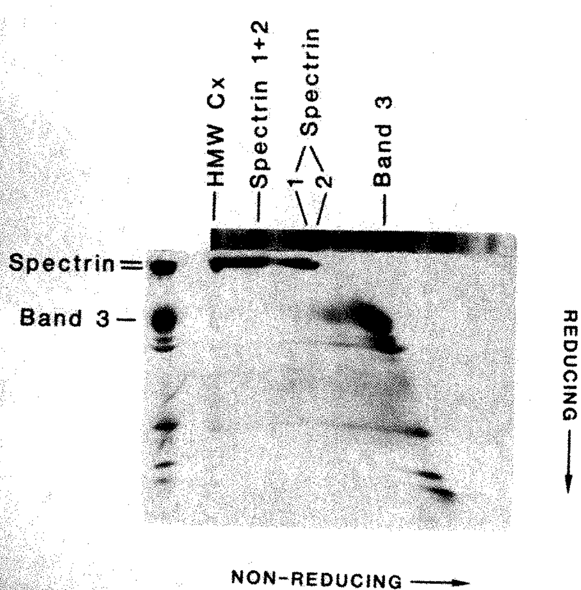


Fig. 2 Two-dimensional SDS-polyacrylamide gel electrophoresis of membranes from diamide-treated erythrocytes. Erythrocytes were diamide-treated, membrane ghosts prepared and membranes were electrophoresed in the first non-reducing dimension as in Fig. 1. This gel was then overlaid on a 7% acrylamide slab gel and electrophoresed in reducing conditions[14].

of band 3 can be controlled by reversible interaction of band 3 with spectrin via band 2.1. Second, Cherry *et al.* proposed that spectrin forms a meshwork underneath the membrane, with interstices through which band 3 can move[12]. Thus, lateral diffusion of transmembrane proteins could be modulated by the state of the spectrin meshwork itself, independently of spectrin-band 3 interactions. The latter hypothesis was supported by observations of Schindler *et al.* that agents which disrupt cytoskeletal structure also free band 3 to diffuse laterally in the membrane[11]. Our results are also consistent with this model of the membrane skeleton as a dynamic structure that modulates lateral movement of membrane proteins. When this dynamic state is stabilized, for example by cross-linking with spectrin, mobility of transmembrane proteins is reduced.

This work was supported by NIH grant HL15157.

Received 29 December 1981; accepted 31 March 1982.

1. Sheetz, M. P., Schindler, M. & Koppel, D. E. *Nature* **285**, 510–512 (1980).
2. Golan, D. E. & Veatch, W. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2537–2541 (1980).
3. Fowler, V. & Bennett, V. *J. supramolec. Struc.* **8**, 215–221 (1978).
4. Haest, C. W. M., Kamp, D., Plasa, G. & Deuticke, B. *Biochim. biophys. Acta* **469**, 226–230 (1977).
5. Haest, C. W. M., Fischer, T. M., Plasa, G. & Deuticke, B. *Blood Cells* **6**, 539–553 (1980).
6. Fischer, T. M., Haest, C. W. M., Stohr, M., Kamp, D. & Deuticke, B. *Biochim. biophys. Acta* **510**, 270–282 (1978).
7. Palek, J. & Liu, S.-C. *Red Cell Proc. 5th int. Conf. on Red Cell Metabolism & Function* (ed. Brewer, G.) (Liss, New York, in the press).
8. Weinstein, R. S., Khodadad, J. K. & Steck, T. L. *J. Cell Biol.* **87**, 209a (1980).
9. Fowler, V. & Branton, D. *Nature* **268**, 23–26 (1977).
10. Huang, H. W. *J. theor. Biol.* **40**, 11–17 (1973).
11. Schindler, M., Koppel, D. E. & Sheetz, M. P. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1457–1461 (1980).
12. Cherry, R. J., Burkli, A., Busslinger, M., Schneider, G. & Parish, G. R. *Nature* **263**, 389–393 (1976).
13. Liu, S. C., Fairbanks, G. & Palek, J. *Biochemistry* **16**, 4066–4074 (1977).
14. Laemmli, U. K. *Nature* **227**, 680 (1970).

Higher plant tubulin identified by self-assembly into microtubules *in vitro*

Louis C. Morejohn & Donald E. Fosket

Department of Developmental and Cell Biology, University of California, Irvine, California 92717, USA

Microtubules are filamentous, subcellular structures present in virtually all eukaryotes. In higher plants, microtubules form the mitotic spindle[1], determine the plane of cell division[2,3] and orient cellulose microfibril deposition in growing cells[4,5], and thus are significant determinants of morphogenesis[6,7]. The main structural component of microtubules is tubulin, a conserved, heterodimeric protein having a molecular weight of 110,000 and composed of α- and β-subunits (≃55,000 each)[8]. The resistance of plant microtubules to several antimitotic drugs has suggested that plant tubulins differ from animal tubulins (see refs 5, 9, 10 for review), but further characterization has awaited the development of a method to isolate plant tubulin. As tubulin has no characteristic enzymatic activity, the positive identification of this protein depends on its ability to self-assemble *in vitro* to form structures which have the morphological features of microtubules. Recent studies[11–17] have used indirect methods to isolate tubulin-like proteins from higher plant cell extracts, but none has conclusively identified tubulin by demonstrating its self-assembly into microtubules. We report here the isolation of tubulin from cultured cells of a higher plant and its unequivocal identification by self-assembly into microtubules *in vitro*. Furthermore, we report peptide mapping data which indicate that although the β-subunits of mammalian and higher plant tubulins have been conserved, the α-subunits have diverged during evolution.

Previous work from our laboratory demonstrated that tubulin self-assembly in extracts of Paul's Scarlet rose (PSR) cells is hampered by low tubulin concentrations, by low molecular weight inhibitors and by endogenous peptidases[17]. We have circumvented these problems by designing a tubulin isolation method that is particularly suited to the physiological characteristics of PSR cells grown in suspension. Early exponential phase (day 3–5) cells were used because during this period the concentrations of cellular proteins are highest[18] and those of potentially harmful polyphenolic compounds lowest[19]. Extracts were prepared in a buffer designed to minimize proteolysis and to enzymatically cleave nucleic acids which, as polyanions, are known to inhibit *in vitro* assembly of tubulin[20]. In addition, anion-exchange chromatography was used to select and concentrate tubulin from extracts by virtue of its conserved polyanionic charge[21].

PSR cells, originally isolated from rose stem tissues (*Rosa* sp. cv. Paul's scarlet) by Nickell and Tulecke[22], were grown as stock suspension cultures in a defined medium[23], modified as previously described[17]. For the propagation of cells to be used for protein isolation, aliquots of 14-day-old stationary phase cells were diluted 6.6-fold with fresh medium and cultured for 3–5 days.

Extracts of cultured cells were prepared, and proteins were fractionated on DEAE-Sephadex A50 as modified from Weisenberg *et al.*[24] (see Fig. 1 legend) and analysed by SDS–polyacrylamide gel electrophoresis[25] (SDS-PAGE). Step gradient elution of the column separated the crude supernatant into three protein-containing fractions designated A, B and C (Fig. 1). SDS-PAGE of these fractions demonstrated that most of the cellular protein eluted in fractions A and B, while fraction C was enriched for two polypeptides with mobilities similar to those of the α- and β-subunits of bovine brain tubulin. Densitometer tracings of stained gels revealed that these bands comprised ~63% of the fraction C polypeptides that entered the gel. Thus DEAE chromatography represents a significant step in the purification of tubulin-like polypeptides from a heterologous population of proteins in PSR cell extracts. The yield of presumptive tubulin obtained by this procedure was 1–2 $\mu g\ mg^{-1}$ of crude supernatant proteins as determined by BioRad assays[26].

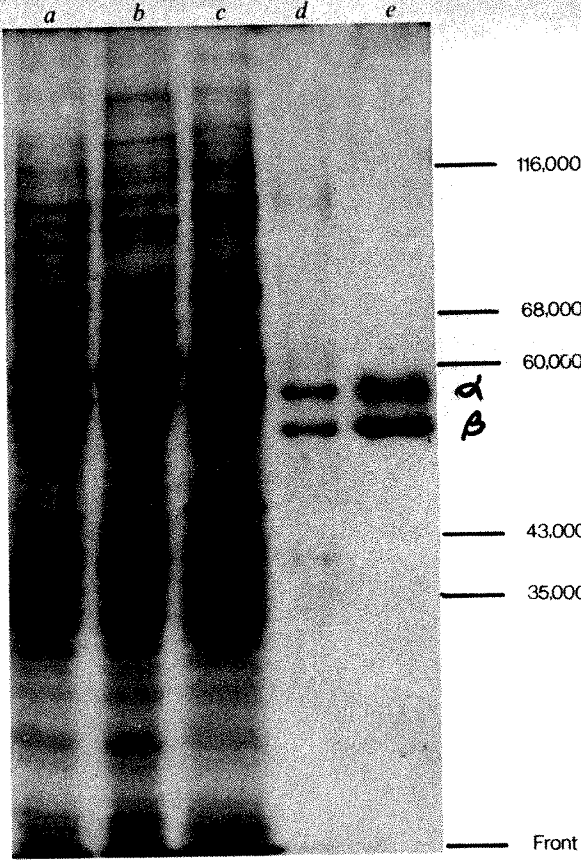


Fig. 1 Analysis of fractionated PSR proteins by SDS-PAGE[25]. Early exponential phase PSR suspension cultures were placed in a dark cold room at 4 °C with constant shaking for 2 h before collection on a Miracloth filter, washed with cold PM buffer (50 mM PIPES-KOH *p*H 6.9, 1 mM EGTA, 0.5 mM $MgCl_2$, 1 mM dithiothreitol). The cells were then weighed and homogenized in a chilled Sorval Omni-mixer in PM buffer containing 2 mM GTP (5 g cells per ml buffer) for 3 min. The homogenate was filtered through three layers of Miracloth and the filtrate was made 25 μM in leupeptin hemisulphate (Sigma) and 100 $\mu g\ ml^{-1}$ in both DNase I (Sigma) and RNase (Calbiochem). The filtered homogenate was centrifuged at 48,200 *g* for 30 min at 4 °C. The crude supernatant, which contained 2.4–3.3 $mg\ ml^{-1}$ of protein as assayed by the method of BioRad[26], was poured into a flask containing DEAE-Sephadex A50 (ref. 24) (pre-equilibrated with PM buffer) at a ratio of 0.5 ml packed volume of resin per ml of crude supernatant. The mixture was swirled in the dark at 4 °C for 1 h, briefly de-aerated and poured into a 2.6 × 40 cm column. The column was packed with crude supernatant at a flow rate of 2–3 ml min^{-1}, and unbound proteins (fraction A) were eluted directly. Weakly bound proteins (fraction B) were eluted with 3–5 bed volumes of PM buffer containing 0.4 M KCl and 0.5 mM GTP. Tightly bound proteins (fraction C) were eluted from the column with 3 bed volumes of PM buffer containing 0.8 M KCl and 0.5 mM GTP. Each fraction was collected by precipitation by 50% saturation with $(NH_4)_2SO_4$ for 30 min. Immediately prior to electrophoresis, $(NH_4)_2SO_4$ precipitates were dissolved in PM buffer, desalted on Sephadex G25 columns and acetone-precipitated, and the precipitates were boiled in SDS-sample buffer[25] for 2 min. Crude supernatant samples were boiled directly in equal volumes of SDS-sample buffer. SDS-derivatives were loaded on to 9.25% slab gels prepared according to Studier[25], except that the stacking gel was 5.1% acrylamide, the resolving gel contained 0.15 M Tris-HCl *p*H 8.8, and the SDS (Sigma) was ethanol-recrystallized. Gels were stained overnight with 0.025% Coomassie brilliant blue R-250, destained, dried and photographed with a yellow filter. *a*, PSR crude supernatant polypeptides; *b*, fraction A polypeptides; *c*, fraction B polypeptides; *d*, fraction C polypeptides containing putative tubulin α- and β-subunits; *e*, three times assembled bovine brain tubulin α- and β-subunits prepared according to ref. 33. Molecular weight standards were *Escherichia coli* β-galactosidase (116,000), bovine serum albumin (68,000), bovine liver catalase (60,000), ovalbumin (43,000) and bovine heart lactate dehydrogenase (35,000).

The proteins in fraction C were tested for their ability to form microtubules *in vitro*. Assembly reactions were performed following the addition of glycerol[27] or taxol[28], an experimental

0028–0836/82/220426—03$01.00

antitumour drug (Fig. 2 legend). Electron microscopy of the materials assembled in the presence of taxol revealed numerous short microtubules, while materials formed in the presence of glycerol contained microtubules which were less abundant and usually longer (Fig. 2). Protein determinations[26] on the microtubule pellets showed that ~7% of the total protein present in fraction C formed pelletable aggregates in the glycerol-treated sample, whereas ~44% of the total protein was pelleted from the taxol-treated sample.

Analysis of the *in vitro*-assembled materials by SDS-PAGE[25] showed that tubulin was the principal protein present in the pelleted aggregates assembled in the presence of both glycerol and taxol (Fig. 3). Furthermore, the rose tubulin polypeptides had mobilities similar to those of bovine brain tubulin polypeptides. In this particular gel system, the β-subunits co-migrated and the rose α-subunit had a slightly faster mobility than that of the brain α-subunit (compare subunits in lanes *d*, *e* of Fig. 3). Similarly, the α-subunits of flagellar tubulins from the green alga *Chlamydomonas* and the ferns *Marselia* and *Pteridium* have been shown to exhibit electrophoretic mobilities different from that of mammalian α-tubulin[29,30].

Densitometer tracings of electrophoretically separated microtubule pellets showed that tubulin represented ~36% of the total protein in the glycerol-assembled material, and ~74% of the taxol–microtubule pellet. Apparently a single round of assembly did not enrich for tubulin in the glycerol-treated samples but did so substantially in the presence of taxol. Horwitz and co-workers have shown that taxol enhances both the assembly rate and yield of brain microtubules formed *in vitro*[28,31]. They have further demonstrated that taxol cannot be readily removed from microtubules, and that it stabilizes the polymers against depolymerization by cold treatment, $CaCl_2$, podophyllotoxin or vinblastine[32]. For these reasons taxol-induced microtubules cannot be further purified by successive rounds of polymerization. Also, the low yield of tubulin from PSR cells and the low efficiency of glycerol-induced microtubule assembly have prohibited the use of cyclic assembly procedures[33]. We are now investigating methods to improve the yield of tubulin from these cells.

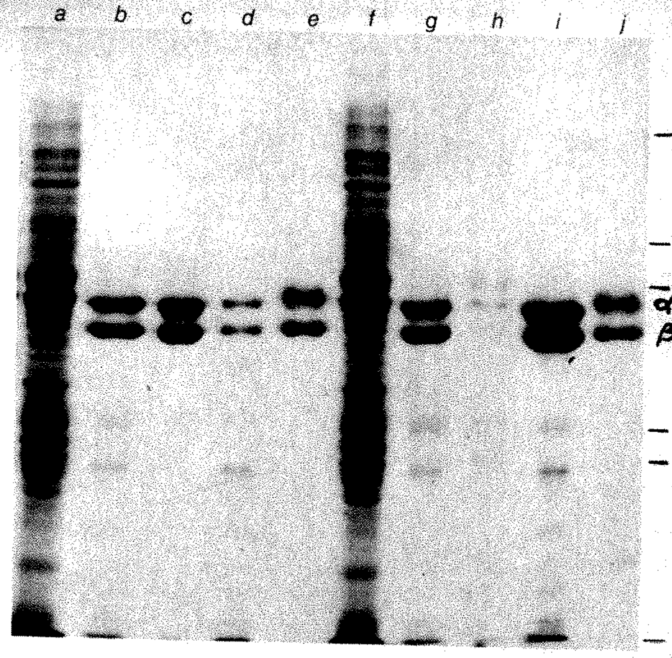


Fig. 3 Analysis of microtubule assembly mixtures by SDS-PAGE[25]. The supernatant and pellet samples described in Fig. 2 legend were boiled in equal volumes of SDS-sample buffer[25] and loaded on to the gel in equal volumes. In this way, we could compare directly the polypeptide composition and abundance of tubulin in each fraction from the glycerol (lanes *c* and *d*) and taxol (lanes *h* and *i*) treatments. *a*, *f*, PSR crude supernatant polypeptides; *b*, *g*, fraction C polypeptides; *c*, glycerol-treated supernatant polypeptides; *d*, glycerol-treated microtubule pellet polypeptides; *e*, *j*, three times assembled bovine brain tubulin[33]; *h*, taxol-treated supernatant polypeptides; *i*, taxol-treated microtubule pellet polypeptides. Molecular weight markers as in Fig. 1.

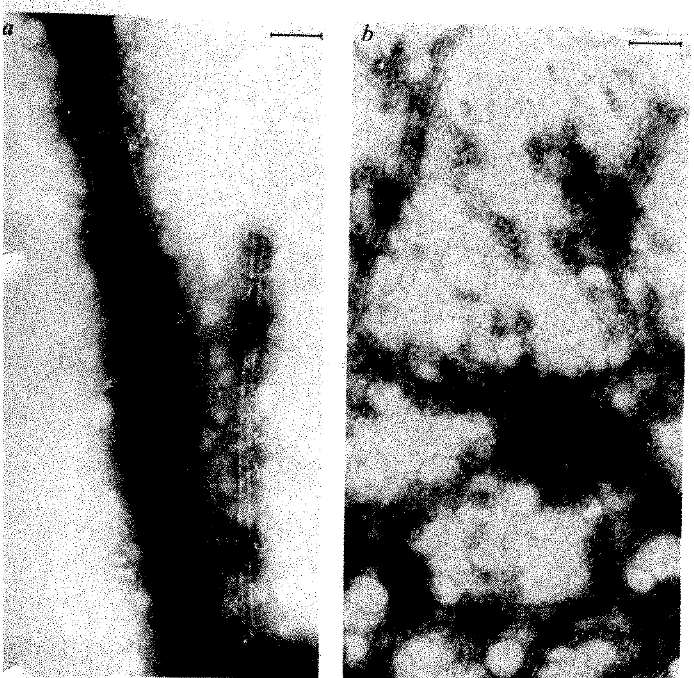


Fig. 2 Electron micrographs of negatively stained rose microtubules assembled *in vitro*. For *in vitro* assembly studies, the fraction C $(NH_4)_2SO_4$ precipitates (described in Fig. 1 legend) from ~200 g of cells were dissolved in 1–1.5 ml of PM buffer containing 2 mM GTP. The solution was desalted on a Sephadex G25 column (5 ml bed volume) which had been pre-equilibrated with the same buffer. The excluded volume was made 50 μM in leupeptin hemisulphate and concentrated by dialysis against 40 ml of the above buffer containing 8 M glycerol and 50 μM leupeptin hemisulphate at 4 °C for 2 h. This reduced the volume of the sample ~10-fold to give a 100–200 μl sample containing 2.6–3.6 mg ml^{-1} of protein as determined by the BioRad[26] procedure. This sample was tested for its ability to form microtubules after addition of glycerol or taxol. The sample was divided equally into two samples, one of which was made 4 M in glycerol[27], while the other was made 10 μM in taxol[28]. The samples were incubated at 37 °C for 1 h to permit microtubule assembly and centrifuged at 130,000 *g* in a Beckman airfuge for 1 h at 23 °C. Each supernatant was removed and divided into two aliquots, one to be used for protein determination and the other for SDS-gels. Each microtubule pellet was resuspended in 10 μl of PM buffer containing 4M glycerol. One μl of the suspension was removed for analysis by electron microscopy, while the remaining suspension was divided into two aliquots, one to be used for protein determinations and the other for SDS-gels. Samples for electron microscopy were placed on carbon- and formvar-coated copper grids, displaced with water, stained with 1% uranyl acetate for 30 s and air-dried. Samples were photographed on a Zeiss 9S electron microscope. *a*, Microtubules assembled in glycerol-treated sample; *b*, microtubules assembled in taxol-treated sample. Bar, 0.1 μm.

Previous studies on ion-exchange-purified mammalian brain tubulin have shown the *in vitro* assembly reaction to be enhanced by both glycerol[27] and taxol[31]. The present study demonstrates that plant tubulin isolated by DEAE chromatography assembles into microtubules after addition of glycerol or taxol. Both the number of microtubules and the yield of tubulin increased when taxol was present. Similar results have been obtained in studies on taxol-induced mammalian tubulin assembly[28,31], and so it seems that the tubulins from organisms as diverse as mammals and higher plants are subject to similar constraints in terms of their abilities to assemble into microtubules *in vitro*.

The microtubules in algae, fungi, ferns and higher plants show resistance to several antimicrotubule agents commonly used to disrupt animal microtubules, suggesting significant differences between plant and animal tubulins[5,10]. Indeed, recent studies have revealed several differences between the tubulins from some of these organisms and from mammals, including variations in electrophoretic mobility[29,34,35], sensitivity to drugs during *in vitro* assembly[35–37] and peptide mapping patterns[30,34].

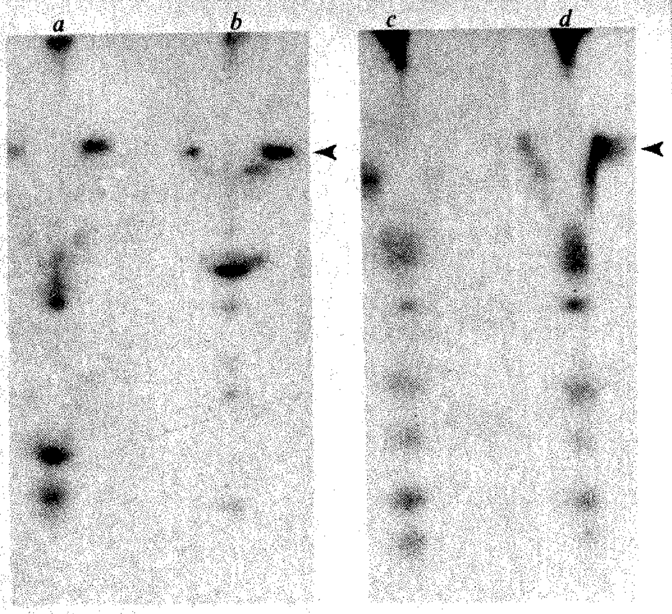


Fig. 4 Peptide mapping analysis of the α- and β-subunits of bovine brain tubulin and rose tubulin. Bovine brain tubulin purified by three rounds of assembly[33] and rose tubulin purified by taxol-induced assembly were analysed by limited proteolysis in gels according to the method of Cleveland *et al.*[38], with minor modifications. The subunits of these tubulins were first separated on the gel system described in Fig. 1 legend. Gels were stained with 0.2% Coomassie brilliant blue R-250 in 50% methanol for 30 min and destained in 5% methanol for 15 min. Bands corresponding to the α- and β-subunits were cut out, equilibrated in a solution of 0.125 M Tris-HCl *p*H 6.8, 0.1% SDS and 1 mM EDTA for 1 h, and loaded on to 15% Laemmli gels[39] together with 0.5 μg of *Staphylococcus aureus* V8 protease. Electrophoresis was performed according to Cleveland *et al.*[38]. Gels were stained and processed as described for Fig. 1. Lanes contain polypeptide fragments from *a*, bovine brain α-subunit; *b*, rose α-subunit; *c*, bovine brain β-subunit; *d*, rose β-subunit. Arrows indicate positions of uncleaved subunits.

We have started structural analysis of cytoplasmic tubulin from a higher plant by preparing peptide maps[38] of the α- and β-subunits of rose tubulin and bovine brain tubulin (Fig. 4). Cleavage of the β-subunits of both tubulins by *Staphylococcus aureus* V8 protease produced very similar patterns of polypeptide fragments. In contrast, the patterns of peptide fragments generated from the cleavage of the α-subunit of both tubulins not only differed markedly from each other, but also from the respective β-subunit patterns. These data suggest that the β-subunits of plants and animals have been more conserved during evolution than the α-subunits.

Recently published[30] peptide maps of the tubulin subunits from cow, squid, sea urchin, echiurid worm, a fern and an alga show the β-subunit maps of all these organisms to be quite similar, the α-subunit maps less similar and the plant α-subunit maps markedly different from those of the animals. Moreover, the fern and algal α-subunit maps resemble the α-subunit map of slime mold tubulin reported by Clayton *et al.*[34]. Our peptide mapping data on higher plant tubulin are consistent with the idea[34] that the α-subunits of tubulins from organisms that are generally resistant to anti-microtubule agents have diverged with respect to the α-subunits from organisms sensitive to those agents. Experiments are in progress to determine whether the resistance of higher plant cells to anti-microtubule drugs is, in fact, due to a low affinity of plant tubulin for these drugs, or to other modes of resistance[10].

We are raising antibodies to plant tubulin subunits for use in further studies on these polypeptides. Efforts are also underway to construct specific cDNA probes for use in studies on the regulation of synthesis of tubulin mRNAs in higher plants.

We thank Dr Susan Horwitz for her generous gift of taxol and Drs Marc Kirchner and Timothy Bradley for helpful suggestions during this investigation. This research was supported by grants to D.E.F. from the NSF (7722398) and the Monsanto Chemical Co., St Louis, Missouri.

Received 5 March; accepted 2 April 1982.

1. Nicklas, R. B. in *Molecules and Cell Movement* (Raven, New York, 1975).
2. Pickett-Heaps, J. D. & Northcote, D. H. *J. Cell Sci.* **1**, 109–120 (1966).
3. Hepler, P. K. & Newcomb, E. H. *J. ultrastruct. Res.* **19**, 498–513 (1967).
4. Hepler, P. K. & Fosket, D. E. *Protoplasma* **72**, 213–236 (1971).
5. Hepler, P. K. & Palevitz, B. A. *A. Rev. Pl. Physiol.* **25**, 309–362 (1974).
6. Gunning, B. E. S. & Hardham, A. R. *Endeavour (Oxf.)* **3**, 112–117 (1979).
7. Hepler, P. K. in *Cytomorphogenesis in Plants* (ed. Kiermayer, O.) (Springer, New York, 1981).
8. Kirchner, M. W. *Int. Rev. Cytol.* **54**, 1–71 (1978).
9. Kihlman, B. A. in *Actions of Chemicals on Dividing Cells* (Prentice-Hall, Englewood Cliffs, 1966).
10. Dustin, P. in *Microtubules* (Springer, Berlin, 1978).
11. Hart, J. W. & Sabnis, D. D. *Planta (Berl.)* **109**, 147–152 (1973).
12. Hart, J. W. & Sabnis, D. D. *J. exp. Bot.* **27**, 1353–1360 (1976).
13. Rubin, R. W. & Cousins, E. H. *Phytochemistry* **15**, 1837–1839 (1976).
14. Slabas, A. R., MacDonald, G. & Lloyd, C. W. *FEBS Lett.* **110**, 77–79 (1980).
15. Yadav, N. S. thesis, Michigan State Univ. (1980).
16. Mizuno, K., Koyama, M. & Shibaoka, H. *J. Biochem.* **89**, 329–332 (1981).
17. Fosket, D. E., Morejohn, L. C. & Westerling, K. E. in *Metabolism and Molecular Activities of Cytokinins* (eds Guern, J. & Péaud-Lenoël, C.) 193–211 (Springer, Berlin, 1981).
18. Nash, D. T. & Davies, M. E. *J. exp. Bot.* **23**, 75–91 (1972).
19. Davies, M. E. *Planta (Berl.)* **104**, 50–65 (1972).
20. Bryan, J., Nagle, B. W. & Doenges, K. H. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3570–3574 (1975).
21. Murphy, D. B. & Borisy, G. G. *Proc. natn. Acad. Sci. U.S.A.* **72**, 2696–2700 (1975).
22. Nickell, L. G. & Tulecke, W. *Bot. Gaz.* **120**, 245–250 (1959).
23. Nesius, K. K., Uchytil, L. E. & Fletcher, J. S. *Planta (Berl.)* **106**, 173–176 (1972).
24. Weisenberg, R. C., Borisy, G. G. & Taylor, E. W. *Biochemistry* **1**, 4466–4479 (1968).
25. Studier, F. W. *J. molec. Biol.* **79**, 237–248 (1973).
26. Bradford, M. M. *Analyt. Biochem.* **72**, 248–254 (1976).
27. Lee, J. C. & Timasheff, S. N. *Biochemistry* **14**, 5183–5187 (1975).
28. Schiff, P. B., Fant, J. & Horwitz, S. B. *Nature* **277**, 665–667 (1979).
29. Ludueña, R. F., Myles, D. G. & Pfeffer, T. A. *Expl Cell Res.* **130**, 455–459 (1980).
30. Little, M., Ludueña, R. F., Langford, G. M., Asnes, C. F. & Farrell, K. *J. molec. Biol.* **149**, 95–107 (1981).
31. Schiff, P. B. & Horwitz, S. B. *Biochemistry* **20**, 3247–3252 (1981).
32. Parness, J. & Horwitz, S. B. *J. Cell Biol.* **91**, 479–487 (1981).
33. Sloboda, R. D., Dentler, W. L. & Rosenbaum, J. L. *Biochemistry* **15**, 4497–4505 (1976).
34. Clayton, L., Quinlan, R. A., Roobol, A., Pogson, C. I. & Gull, K. *FEBS Lett.* **115**, 301–305 (1980).
35. Kilmartin, J. V. *Biochemistry* **20**, 3629–3633 (1981).
36. Roobol, A., Pogson, C. I. & Gull, K. *Expl Cell Res.* **130**, 203–215 (1980).
37. Quinlan, R. A., Roobol, A., Pogson, C. I. & Gull, K. *J. gen. Microbiol.* **122**, 1–6 (1981).
38. Cleveland, D. W., Fischer, S. G., Kirchner, M. W. & Laemmli, U. K. *J. biol. Chem.* **252**, 1102–1106 (1977).
39. Laemmli, U. K. *Nature* **227**, 680–685 (1970).

Genes for immunoglobulin heavy chains and for α_1-antitrypsin are localized to specific regions of chromosome 14q

Diane W. Cox, Vera D. Markovic* & Ikuko E. Teshima

The Hospital for Sick Children, 555 University Avenue, Toronto, Canada, and Departments of Paediatrics and Medical Genetics, University of Toronto, Toronto, Ontario M5G 1X8, Canada and
* Genetics Department, Surrey Place Centre, Toronto M5S 2C2, Canada

Studies of somatic cell hybrids have assigned the gene cluster for human immunoglobulin heavy chains μ, γ and α to chromosome 14 (ref. 1). The locus for Gm (*GM*), a genetic marker on the γ heavy chain, has been shown to be linked to *PI*, the locus for α_1-antitrypsin/α_1-protease inhibitor (α_1AT), a major protease inhibitor in human serum[2,3]. This assignment has now been confirmed by studies of somatic cell hybrids which have assigned α_1AT also to chromosome 14 (refs 4, 23). From studies of two families having abnormalities involving the long arm of chromosome 14, we have localized *GM* to the terminal portion of 14q at band q32.3 and *PI* to a more proximal position, between q24.3 and q32.1. The immunoglobulin genes are within a chromosome region noted for its high frequency of breaks associated with chromosome rearrangement, which occur both spontaneously in cultured lymphocytes and in certain malignancies.

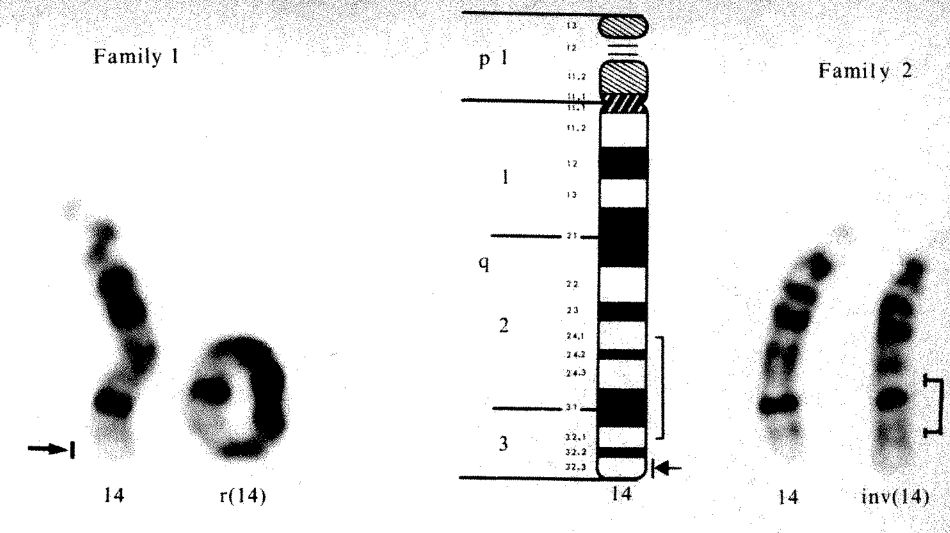


Fig. 1 Chromosomes observed with G-banding. The normal chromosome 14 and r(14) in proband of family 1 is shown with the deleted region marked by an arrow. The normal chromosome 14 and inv(14) in family 2 is shown with breakpoints indicated by brackets. The diagram (centre) shows the prometaphase band pattern according to International Nomenclature[22].

Early metaphase chromosomes in cultured lymphocytes were studied by trypsin–Giemsa G-banding[5]. Twenty or more cells of each individual were examined, at least 10 of which were banded. To define the breakpoint of the ring chromosome, C-banding[6], Giemsa-11[7] and NOR staining[8] were also used. PI typing was carried out by isoelectric focusing in polyacrylamide *p*H 3.5–6, and α_1AT was quantitated by electroimmunoassay[9]. GM allotypes were determined by haemagglutination inhibition using fresh O Rh$^+$ (Rhesus-positive) human erythrocytes sensitized with non-agglutinating anti-Rh antibody[10]. All sera were tested at a 1/20 dilution, with antibody for the γ1 markers *G1m* (*a*, *x* and *f*) and for the γ3 markers *G3m* (*b* and *g*) (reagents were given by Dr Moses Schanfield, American Red Cross). Quantitation of immunoglobulins (IgG, IgM and IgA) was carried out by laser nephelometry (Beckman). Other genetic markers were typed by standard techniques and used to verify the stated family relations: blood groups ABO, Rh and MNS; complement 3 (C3) and properdin factor B (BF) were identified by agarose electrophoresis; and group-specific component (GC) and transferrin (TF) by isoelectric focusing in polyacrylamide gels.

In family 1 (HSC–C4), the proband was an 8-year-old male with mental retardation, seizures, microcephaly and primary pulmonary hypertension. His karyotype, examined from 21 cells the first time and 100 cells on a second occasion, was 46 XY, ring 14 [r(14)]. In the two analyses, 93% of blood lymphocytes had the karyotype 46 XY, r(14), with at least one ring present, while the remainder had one copy of chromosome 14 lacking the ring. Breakpoints were distal to bands p11 and q32.2. The deleted portion of the chromosome is therefore all or part of q32.3 (Fig. 1). Further clinical and cytogenetic details of this patient will be reported elsewhere (V. D. M., R. G. Worton, C. B. Verellen and J. M. Berg, in preparation).

Both parents had normal karyotypes. Results of *PI* and *GM* typing, determined on two separate occasions, are shown in Fig. 2. The proband expressed only the paternal GM type, *fb*, and lacked a maternal GM component. This was confirmed in a repeat blood sample from proband and mother; both sera were also tested at 1/2 dilution. This result can be explained if the GM marker (γ-chain locus) is in the deleted region, q32.3. Inherited deletions of γ1 or γ3 regions have rarely been reported, however the GM types of I-2 and II-1 (Fig. 2) exclude the possibility of an inherited deletion. Serum concentrations of immunoglobulins are shown in Table 1. The IgG, IgA and IgM concentrations in serum of the proband are normal and do not show a dosage effect. This may, however, be the result of allelic exclusion, by which only one of the alleles for the immunoglobulin chains is expressed in any one immunoglobulin-producing cell, suggesting that the intact IgG region is always expressed when there is a deletion in the homologous chromosome. Genetic types of α_1AT (PI types) were not informative as proband and parents were all of PI type M1. Serum concentrations of α_1AT did not suggest a deletion for this locus; heterozygotes for the null allele of α_1AT (*PIM*−) have lower concentrations (59 ± 6.8% of normal)[11] than those found in the proband. Genetic markers were compatible with the stated parentage.

Family 2 (HSC-C3) provided information on localization of the *PI* locus. The proband was a 7-year-old boy with neutropenia. An altered banding pattern of chromosome 14 was observed in all 26 cultured lymphocytes examined by trypsin–Giemsa G-banding (Fig. 1). We concluded that the altered pattern was due to a paracentric inversion with breakpoints at bands q24.1 and q32.1; karyotype 46 XY, inv(14) (q24.1 q32.1). The inverted chromosome 14 was also found in other relatives (Fig. 2). Genetic markers were compatible with the stated family relations. Results of typing are shown in Fig. 2. If the *PI* allele, *M2* and the inversion are assumed to be coupled in I-2, then there are no recombinants among 11 offspring in this family. These results are compatible with location of the *PI* locus at the inversion, between bands q24.1 and q32.1. Genes within the inversion would be inherited together because a single cross-over in the inverted region would result in non-viable gametes. As there is an average of 1.76 chiasmata for chromosome 14q (ref. 12), and if a chiasma is equivalent to a cross-over, then a double cross-over, which would be necessary to produce viable gametes, is not likely to occur in the small distance covered by this inversion. Linkage analysis by lod scores supports the location within the inverted region. The inverted region was considered as a marker. Lod scores for the three sibships, two of which are phase known, were obtained from the tables of C. A. B. Smith as given by Emery[13]. Calculations were simplified by assuming I-1 to be of PI type M1M1, where the relative probability for this PI type, based on population frequencies from our laboratory and from PI types of the offspring, is 87%. The maximum likelihood

Table 1 Serum concentration of immunoglobulins and α_1-antitrypsin in family 1

Protein	Proband	Father	Mother	Controls [mean (95%)] confidence interval
IgG (mg dl^{-1})	808	906	998	915 (633–1280*)
IgA (mg dl^{-1})	54	134	97	90 (33–202*)
IgM (mg dl^{-1})	86	154	186	107 (48–207*)
α_1AT (% normal)	79	84	100	105 (67–198†)

* Values for children 6–8 y old (from C. R. Jolliff and K. Cost); adult values are higher.

† Obtained from 112 normal adults, in our laboratory. Values for children are similar.

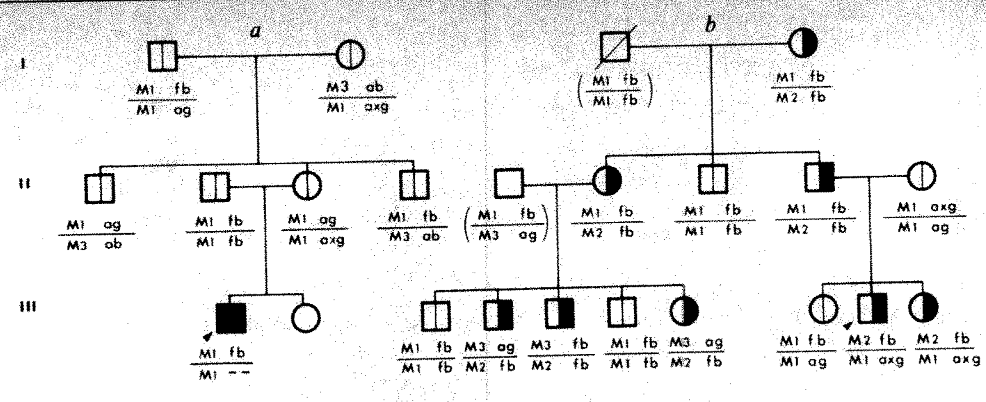


Fig. 2 Pedigrees of family 1 (*a*) and family 2 (*b*) showing abnormal chromosomes and the most probable arrangement of PI and GM types. Arrowheads indicate pro-bands. ■ Indicates the presence of r(14); ○, □ indicate the presence of inv(14).

estimate of recombination (B) from plotted values[12] is 0 with a lod score of 3.0. This indicates there is a high probability (1,000:1) of absolute linkage between *PI* and the inversion marker. A map distance of up to 5 centimorgans (cM) (recombination value, $\theta = 0.05$) would not be excluded.

Our studies in family 1 indicate that *GM* is localized at 14q32.3. While we consider the loss of the structural γ-chain locus to be probable, we cannot exclude the possibility that a promoter or sequences necessary for initiation of the γ-chain switch have been lost. Studies of the DNA of the patient carrying the deletion might clarify this question. We have recently learned, however, that *in situ* hybridization studies in the laboratory of P. Leder, using heavy-chain constant-region probes, also confirm that γ-chain loci are localized at 14q32, a region larger than, and including, our region 14q32.3 (I. Kirsch, personal communication).

Data from family 2 indicate that the *PI* locus is in the region q24.1–q32.1. The *GM–PI* linkage in males is 0.26[3] for 'non-Z' alleles. As there is an average of 1.76 chiasmata for chromosome 14q (ref. 12), and a recombination value of 0.5 in males is defined as equivalent to 50 cM, the genetic length of 14q is ~88 cM in human males. The recombination value of 0.26 for *GM–PI* falls either at the distal border of q24.3, according to Francke's band patterns based on actual chromosome measurements[14] or at the proximal portion of q24.3 in Fig. 1. However, the distribution of chiasmata in 14q is uneven, with a high concentration in terminal portions of the chromosome, thus the most likely location for *PI* is in a more distal region of the inversion. As it is unknown how the genetic map corresponds to the physical map, we must localize *PI* at q24.3–32.1. Nucleoside phosphorylase, which has been localized at the region q12–13 (ref. 15), should not be linked with *GM*, but may show linkage with *PI*, the minimal distance estimated as being 24 cM.

Our localization of *GM* at band q32.3 may have important biological implications as it is within or close to a region of the chromosome that is prone to breakage. Analysis of spontaneous breaks in cultured lymphocytes has indicated that chromosome 14 was involved in rearrangements more often than any other chromosome[16]. The 'hotspots' for breaks were 14q13 and 14q32. In ataxia telangiectasia, an inherited condition in which chromosome breaks occur frequently in cultured cells, the chromosome abnormalities frequently involve the long arm of chromosome 14, sometimes involving a tandem duplication with breaks through q11 and q32 (for review see ref. 17). In Burkitt's lymphoma and in other lymphomas, chromosome breaks are non-random. Burkett's lymphoma clones frequently contain rearrangements between chromosomes 8 and 14, specifically t(8; 14) (q24; q32)[17]. Other abnormalities of chromosome 14 are found with high frequency, consistently with breaks at q32. DNA rearrangements, which occur in the immunoglobulin gene cluster both during the switch from one type of heavy chain to the other and within differentiated cells to produce antibody variability (reviewed in ref. 18), may be related to this high frequency of breakage. The genes for human light-chain (λ) immunoglobulin chains have been assigned to chromosome 22 and association of this chromosome with translocations in malignancies has been noted[19]. The λ gene is also a region which undergoes DNA rearrangement. Malignant clone proliferation may be related to effects on the immunoglobulins, if breaks occur directly within the gene.

A restriction enzyme fragment of human DNA cloned in phage λ CH4-rHs18 has been found to be highly polymorphic, apparently because of DNA rearrangements rather than base pair substitutions or modifications[20]. This locus, D14S1, has recently been localized at the distal half of 14q[21]. If regions of DNA rearrangement do indeed promote chromosome instability, then it will be of considerable interest to determine whether this locus is the same as, or close to, the immunoglobulin locus.

We thank Dr Joseph Berg who originally made us aware of family 1; Dr Ron Worton for helpful discussions; Drs Moses Schanfield and Linda Walsh for helping to establish GM typing in our laboratory; Patricia Zavitz for obtaining blood samples from family members; and Dr Louis Siminovitch for review of the manuscript. We also acknowledge the technical assistance of Tammy Mansfield, Gail Billingsley, Shirley Smyth, Diane Wills, Vera Foldes, Aileen Brown, Chin Ho, Hannah Ho, Kim Mellick and Kwan Ng. This study was supported in part by a grant from the MRC of Canada (MA5426). I.E.T. was supported by The Hospital for Sick Children, in the laboratory of R. Worton.

Received 7 January; accepted 23 March 1982.

1. Croce, C. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **76,** 3416–3419 (1979).
2. Gedde-Dahl T. Jr, Cook, P. J. L., Fagerhol, M. K. & Pierce, J. A. *Ann. hum. Genet.* **39,** 43–50 (1975).
3. Gedde-Dahl, T. Jr, *et al. Ann. hum. Genet.* **45,** 143–153 (1981).
4. Pearson, S., Tetri, P., George, D. L. & Francke, U. *Am. J. hum. Genet.* **33,** 148A (1981).
5. Seabright, M. *Lancet* **ii,** 971–972 (1971).
6. Arrighi, F. E. & Hsu, T. C. *Cytogenetics* **10,** 81–86 (1971).
7. Bobrow, M., Madan, K. & Pearson, P. L. *Nature new Biol.* **238,** 122–124 (1972).
8. Bloom, S. E. & Goodpasture, C. *Hum. Genet.* **34,** 199–206 (1976).
9. Cox, D. W. *Am. J. hum. Genet.* **33,**354–365 (1981).
10. Schanfield, M. S., Polesky, H. F. & Sebring, E. B. in *Paternity Testing* (ed. Polesky, H. F.) 45–54 (ASCP-EMS, Chicago, 1975).
11. Cox, D. W., Billingsley, G. D. & Smyth, S. in *Proc. 'Electrophoresis 81'*, 505–510 (Verlag-Chemie, Germany).
12. Hulten, M. *Hereditas* **76,** 55–78 (1974).
13. Emery, A. E. H. in *Methodology in Medical Genetics*, 63–75 (Churchill-Livingstone, Edinburgh, 1976).
14. Francke, U. *Cytogenet. Cell Genet.* **31,** 24–32 (1981).
15. George, D. L. & Francke, U. *Science* **194,** 851–852 (1976).
16. Mattei, M. G., Ayme, S., Mattei, J. F., Aurran, Y. & Giraud, F. *Cytogenet. Cell Genet.* **23,** 95–102 (1979).
17. Sandberg, A. in *The Chromosomes in Human Cancer*, 161–165, 400–405 (Elsevier/North Holland, New York, 1980).
18. Marx, J. L. *Science* **212,** 1015–1017 (1981).
19. Erikson, J., Martinis, J. & Croce, C. M. *Nature* **294,** 173–175 (1981).
20. Wyman, A. R. & White, R. *Proc. natn. Acad. Sci. U.S.A.* **77,** 6754–6758 (1980).
21. de Martinville, B., Wyman, A. R., White, R. & Francke, U. *Am. J. hum. Genet.* **34,** (in the press).
22. Harnden, D. G. *et al.* (eds) *Cytogenet. Cell Genet.* **31,** 1–23 (1981).
23. Darlington, G. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **79,** 870–873 (1982).

Isolation and structural organization of the human preproenkephalin gene

Masaharu Noda, Yutaka Teranishi, Hideo Takahashi, Mitsuyoshi Toyosato, Mitsue Notake, Shigetada Nakanishi* & Shosaku Numa

Department of Medical Chemistry and * Institute for Immunology, Kyoto University Faculty of Medicine, Kyoto 606, Japan

Recently, we have elucidated the primary structure of bovine adrenal preproenkephalin by determining the nucleotide sequence of cloned DNA complementary to its mRNA[1]. The structure of most of this precursor molecule has also been deduced by Gubler *et al.*[2] using cDNA sequencing in conjunction with protein sequencing. Bovine preproenkephalin contains four copies of methionine-enkephalin[3] (Met-enkephalin) and one copy each of leucine-enkephalin[3] (Leu-enkephalin), Met-enkephalin-Arg^6-Phe^7 (ref. 4) and Met-enkephalin-Arg^6-Gly^7-Leu^8 (refs 1, 2, 5). The region containing the repeated enkephalin and extended enkephalin sequences, which are each bounded by paired basic amino acid residues, is connected with a cysteine-containing amino-terminal sequence preceded by a signal peptide[6]. We have now studied the relationship between the repetitive structure of preproenkephalin and the structural organization of its gene by cloning a human genomic DNA segment containing the entire gene. We find that the general organization of the preproenkephalin gene is strikingly similar to that of the gene encoding the common precursor of corticotropin (ACTH) and β-lipotropin (β-LPH)[7–9] (alternatively designated preproopiomelanocortin), another multi-hormone precursor. Furthermore, the complete mRNA and amino acid sequences of human preproenkephalin have been deduced from the corresponding gene sequence.

A human genomic DNA library[10], kindly provided by Dr T. Maniatis, was screened for phage carrying preproenkephalin gene sequences by hybridization *in situ*[11] at 60 °C with the bovine cDNA probe[1]. The DNA inserts carried by the isolated hybridization-positive phage clones were subjected to restriction mapping and blot hybridization analysis[12] with the cDNA probe (see Fig. 1 legend for details of cloning). Figure 1*a* shows a restriction map thus obtained for the hybridization-positive 2.0- and 6.7-kilobase pair (kbp) *Eco*RI fragments, which were shown to be adjacent to each other. This restriction map is in agreement with that of the mRNA-coding sequence on human placental cellular DNA constructed by blot hybridization analysis with the bovine cDNA and the 2.0-kbp *Eco*RI fragment as probes, thus assuring that the cloned 2.0- and 6.7-kbp *Eco*RI fragments are derived from a single segment of the human genome. (The *Eco*RI site at the 5′ end of the 2.0-kbp fragment is derived from the *Eco*RI linker joining this fragment to the phage vector.) These two *Eco*RI fragments were subcloned in the plasmid pBR322 (ref. 13) for further analysis.

Detailed restriction maps for the mRNA-coding sequences (exons) in the 2.0- and 6.7-kbp *Eco*RI fragments were constructed, and the DNA fragments containing an exonic sequence and its surrounding regions were subjected to nucleotide sequence analysis by the procedure of Maxam and Gilbert[14] according to the strategy indicated (Fig. 1*b*, *c*). Comparison of the human DNA sequence determined (Fig. 2) with the nucleotide sequence of the bovine cDNA[1] has enabled us to locate two intervening sequences (introns) in the human preproenkephalin gene. One of them, 469 base pairs (bp) long (intron A), is in the segment encoding the 5′-untranslated region of the mRNA 3 bp upstream from the translational initiation site. The other, ~3.4 kbp long (intron B), interrupts the protein-coding sequence between the triplets encoding amino acids 46 and 47 of preproenkephalin. The segment specifying amino acids 47–267 and the 317-bp segment corresponding to the 3′-untranslated region of the mRNA are uninterrupted. This assignment of the introns is supported by the general rule for exon–intron boundaries according to which an intron begins with a G-T and ends with an A-G dinucleotide[15]. Furthermore, the message strands at the ends of both introns are complementary to the 5′-terminal sequence of U1 small nuclear RNA, as observed for the introns of other genes[16,17].

The cloned bovine cDNA used lacked a sequence of ~65 bp encoding the 5′-terminal region of the mRNA[1]. To identify the gene sequence of this region and the 5′ terminus of the mRNA, we used two methods. First, a 5′-end-labelled DNA probe hybridized to the 5′-untranslated region of the mRNA was treated with S_1 nuclease, and the length of the protected DNA fragment measured by electrophoresing it beside a size marker prepared by chemical cleavage of a sample of the labelled probe[18] (Fig. 3*a*). Second, we estimated the size of the cDNA transcript formed by reverse transcriptase-mediated elongation of a 5′-end-labelled DNA primer hybridized to a segment of the 5′-untranslated region of the mRNA (Fig. 3*b*). The S_1 mapping procedure yielded a single DNA fragment which extended to the position corresponding to the C residue on the message strand located 217 nucleotides upstream from the 5′ end of intron A (see ref. 19 for this positioning). This position agreed approximately with that of the 5′ terminus of the mRNA

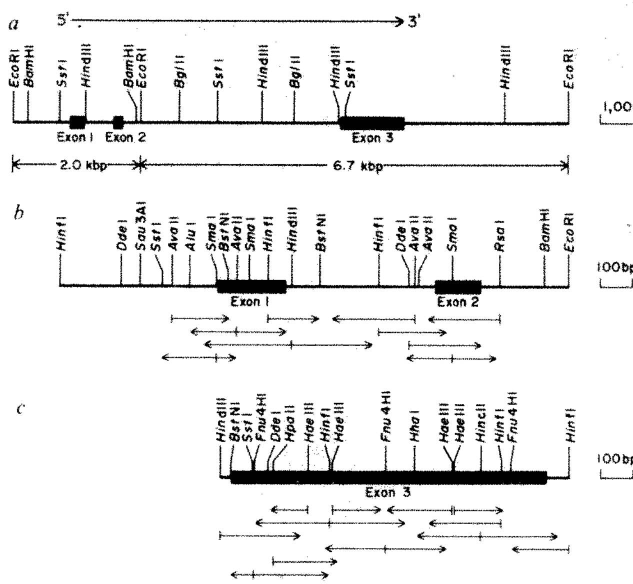


Fig. 1 Restriction mapping of cloned human genomic DNA containing the preproenkephalin gene and sequencing strategy. The library screened was a collection of recombinant phage that carried human fetal liver DNA fragments generated by partial digestion with *Hae*III and *Alu*I and joined to the λ Charon 4A arms with *Eco*RI linkers. The hybridization probe was a mixture of the *Pst*I-cleaved inserts of plasmids pENK5 and pENK45 (ref. 1) labelled with ^{32}P by nick-translation[31]. Five hybridization-positive phage clones were isolated from about 10^6 plaques by repeated plaque purification. Two of these clones (first type) carried a DNA insert containing six *Eco*RI fragments with approximate lengths of 1.1, 1.2, 1.4, 2.0, 2.9 and 6.7 kbp, and the 2.0- and 6.7-kbp fragments hybridized with the cDNA probe[12]. The inserts of the remaining three clones (second type) contained five *Eco*RI fragments approximately 1.1, 1.2, 1.8, 2.9 and 3.2 kbp long, and only the 3.2-kbp fragment was hybridization-positive. The presence of common restriction sites suggested that the two types of clone carried an overlapping DNA segment and that the 3.2-kbp fragment in the clones of the second type constituted a portion of the 6.7-kbp fragment in the clones of the first type. Restriction maps with scales on the right side are given for the DNA segment composed of the 2.0- and 6.7-kbp *Eco*RI fragments (*a*), a portion of the 2.0-kbp *Eco*RI fragment containing exons 1 and 2 (*b*) and a portion of the 6.7-kbp *Eco*RI fragment containing exon 3 (*c*). The direction of transcription is from left to right. For reference, the locations of the exons are shown by closed boxes. All existing sites for the endonucleases indicated are displayed in *a*; only relevant restriction sites are shown in *b* and *c*. The horizontal arrows beneath the restriction maps in *b* and *c* indicate the direction and extent of sequence determinations. Further experimental details have been described previously[8,25].

0028-0836/82/220431—04$01.00

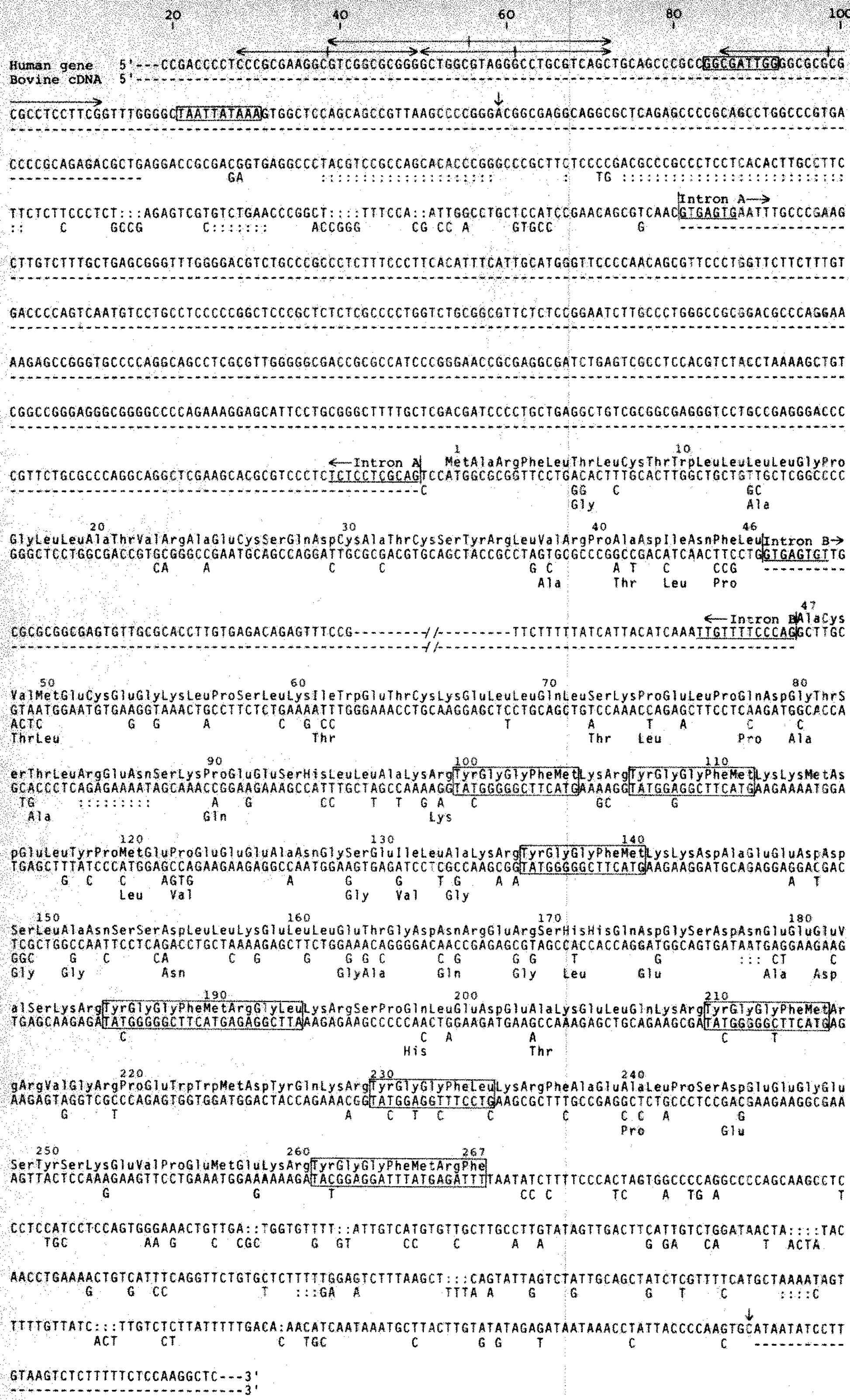


Fig. 2 Nucleotide sequence of the human preproenkephalin gene and comparison with that of the bovine cDNA. The nucleotide sequence of the human message strand, together with the deduced amino acid sequence, is shown, and the nucleotide and amino acid differences found in the bovine sequences[1] (derived from the nucleotide sequences of the cDNA clones pENK5 and pENK45) are displayed beneath the human sequences; the absence of a nucleotide or an amino acid in the bovine sequence indicates that the human and bovine sequences are the same; the presence of a colon in either nucleotide sequence indicates a gap; the hyphens indicate sequences that have not been determined or cloned. The amino acid residues in the human sequence are numbered, beginning with the initiative methionine. The sequences of Met-enkephalin, Leu-enkephalin and Met-enkephalin with a carboxyl extension, which are each bounded by paired basic amino acid residues (or by the carboxy-terminus of preproenkephalin), are boxed together with the coding nucleotides. The exon–intron junctions, which are positioned according to the G-T/A-G rule[15], are indicated by vertical lines. The sequences at both ends of the introns that are complementary to the 5′-terminal sequences of U1 small nuclear RNA are underlined. The putative sites of capping and poly(A) addition are indicated by vertical arrows. The Hogness box and a variant of the model sequence $GG^{C}_{T}CAATCT$ in the 5′-flanking region are boxed. The palindromes in the 5′-flanking region are shown by pairs of horizontal arrows starting at their centre above the nucleotide sequence.

indicated by the primer elongation experiment, so that the 5′-terminal gene sequence is uninterrupted. It is known that mouse β-globin mRNA, analysed by S_1 mapping, yields two protected DNA fragments which are four and five nucleotides longer than expected, perhaps because a cap structure at the 5′ terminus of the mRNA hinders digestion of the DNA probe by S_1 nuclease[20]. Furthermore, most eukaryotic mRNAs start with an A residue[21]. Therefore, we tentatively locate the capping site of the preproenkephalin gene at the A residue 213 bp upstream from the 5′ end of intron A. Consistent with this assignment is the presence of a 'Hogness box' (ref. 22; reviewed in ref. 21) and a variant of the model sequence $GG^{C}_{T}CAATCT$ (ref. 23) 29 and 67 bp, respectively, upstream from the capping site.

On the basis of the results described above, we conclude that the human preproenkephalin gene is ~5.2 kbp long and consists of three exons (designated exons 1, 2 and 3 in the 5′ to 3′ direction) separated by two introns. If the G-T/A-G rule for exon–intron boundaries[15] is applied and if the redundant A residue at the poly(A) addition site is not counted, exon 1 is composed of 213 bp, exon 2 of 141 bp and exon 3 of 980 bp.

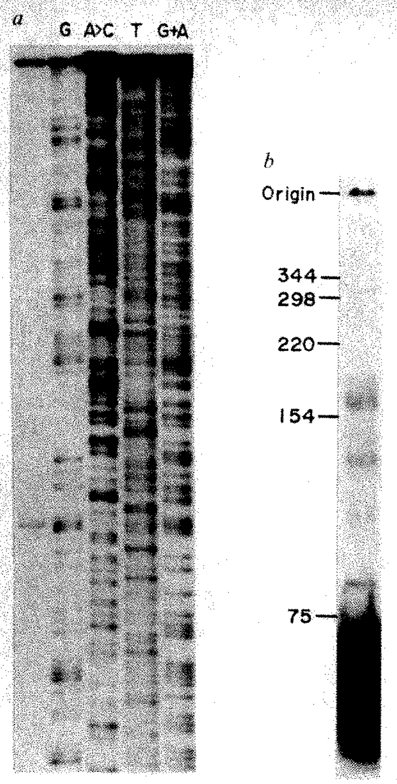


Fig. 3 Identification of the 5′ terminus of human preproenkephalin mRNA by S_1 mapping (*a*) and primer elongation (*b*). Poly(A)-containing RNA was isolated from human pheochromocytomas obtained by tumour resection according to the procedure used previously for the isolation of rat liver RNA[32]. The 2.0-kbp *Eco*RI fragment, isolated from the plasmid carrying it, was cleaved by *Hin*fI, and the ~650-bp *Hin*fI–*Hin*fI fragment (located most upstream in Fig. 1*b*) was isolated and labelled with ^{32}P at the 5′ end. In *a*, this labelled fragment was cleaved by *Sau*3AI, and the ~400-bp *Hin*fI–*Sau*3AI fragment (see Fig. 1*b*) was isolated and used as a probe. The probe (0.03 pmol, 1.2×10^5 c.p.m.) was denatured, hybridized to poly(A) RNA (20 μg) in 80% formamide at 63 °C for 3 h and digested with S_1 nuclease[18]. A sample of the labelled probe was subjected to the G, A>C, T and G+A degradations[14,33] and used as a size marker. The product of S_1 digestion and the size marker were electrophoresed on 7 M urea/8% polyacrylamide gel. In *b*, the 5′-end-labelled ~650-bp *Hin*fI–*Hin*fI fragment was cleaved by *Sma*I, and the resulting products were denatured and electrophoresed on 7 M urea/6% polyacrylamide gel for strand separation[14]. The anti-message strand (62 nucleotides) of the smaller *Hin*fI–*Sma*I fragment (4 pmol, 1.6×10^7 c.p.m.) (see Fig. 1*b*) was hybridized to poly(A) RNA (20 μg) in 0.4 M KCl at 60 °C for 1 h and then subjected to reverse transcriptase reaction[34]. The cDNA formed was electrophoresed on 7 M urea/6% polyacrylamide gel. The size of the largest cDNA species was estimated to be ~160 nucleotides by comparison with the 5′-end-labelled *Hin*fI cleavage products of pBR322 DNA, the sizes of which are given in nucleotides. For further experimental details, see ref. 25.

The structural organization of the preproenkephalin gene, together with the degree of nucleotide sequence homology between the human and bovine protein-coding regions, is schematized in Fig. 4. There are four amino acid additions in the human sequence (Arg-Glu-Asn at positions 85–87 and Asn at position 178). Thus human preproenkephalin consists of 267 amino acids and has a calculated molecular weight of 30,785. The coding regions for the enkephalin and extended enkephalin sequences and the paired basic amino acid residues flanking them exhibit no nucleotide substitution resulting in an amino acid replacement (replacement substitution), except for the substitution of Arg at codon 99 of the human sequence for Lys in the bovine sequence. The coding region for the peptide fragment positioned between the Leu-enkephalin and the preceding Met-enkephalin sequences also shows no replacement substitution. On the other hand, the remaining regions are subjected to some amino acid replacements, although they are fairly well conserved.

We have previously noted that the structural organization of preproenkephalin resembles that of the ACTH–β-LPH precursor in that both proteins are composed of multiple repetitive units (enkephalin or melanotropin sequences) connected with a cysteine-containing amino-terminal sequence preceded by a signal peptide[1]. The two proteins are composed of an identical (human) or almost identical number (bovine) of amino acid residues. The present study reveals a striking similarity between the structural organizations of the genes encoding the two proteins. Both genes contain two introns; one of them is inserted in the segment encoding the 5′-untranslated region of the mRNA near the translational initiation site, and the other in the protein-coding sequence near the signal peptide region at an almost equivalent position.

A characteristic of both the preproenkephalin gene and the ACTH–β-LPH precursor gene is that all the repeated enkephalin or melanotropin sequences are encoded by a single large exon (exon 3). When a new gene like an immunoglobulin heavy chain gene was created by duplication of a primordial DNA segment encoding a functional translation unit together with its flanking regions, the organism might have adopted a pre-existing RNA splicing mechanism to eliminate a spacer between the duplicated units[24]. Unlike the domains of immunoglobulins, the repeated peptide units of the multihormone precursors do not function as parts of an integrated protein molecule, but are destined to be separated from one another by proteolytic cleavage to exhibit coordinated biological activities[1,25,26]. In such a case, the organism might have resorted to a pre-existing protein-processing mechanism to remove the spacer region. The signals specifying the sites of proteolytic cleavage might have existed in the ancestral DNA segment or might have been created *de novo* by drift at the proper positions after duplication. This duplication process could be repeated to produce the characteristic structure of the

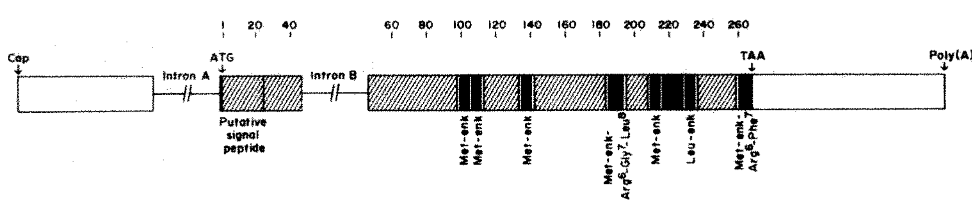


Fig. 4 Structural organization of the human preproenkephalin gene and nucleotide sequence homology between its protein-coding region and the bovine cDNA. The exons are shown by blocks, and the introns by lines; the lengths of the lines do not represent the real lengths of the introns. The sites of capping, translational initiation (ATG), translational termination (TAA) and poly(A) addition are indicated. The segments encoding the 5′- and 3′-untranslated regions of the mRNA are shown by open boxes. Amino acid numbers (see Fig. 2) are given above the gene sequence. The open slits in the protein-coding segment represent the coding nucleotides for the paired basic amino acid residues separating the component peptides, which are displayed beneath the gene sequence. The protein-coding segment is divided into 14 regions bounded by the paired basic residues (or by one of the termini or by the cleavage site of the signal peptide). The regions with and without replacement substitutions are shown by shaded and solid boxes, respectively. The corrected per cent divergence[27] calculated for the replacement substitution sites of each region is as follows (codons corresponding to gaps were excluded from the calculation): signal peptide region (amino acid residues 1–24), 8.5; residues 25–97, 10.5; residues 114–133, 13.9; residues 143–183, 16.4; residues 196–207, 9.4; residues 237–258, 5.9. enk, Enkephalin.

multi-hormone precursors. Because such a protein-processing mechanism could overcome the possible deleterious effect of an inexact deletion of an intron, introns which existed originally between the coding regions for the repeated units of the multi-hormone precursors may have been lost, as proposed for the C-peptide-coding region of the rat preproinsulin I gene[27,28]. Once a flanking peptide, connected with a duplicated unit, has acquired a specific function[1,25,26], however, the deletion of an intron would have to be precise, and this would be a rather rare event. In view of the fact that the multi-hormone precursor genes contain, in addition to the coding regions for the repeated units, many segments that apparently represent direct duplications[1,25], it seems more likely that the repetitive structure of these genes has evolved by duplication of an ancestral DNA segment at adjacent positions in conjunction with a protein-processing mechanism. On the other hand, the two introns of these genes, located upstream of the region encoding the multiple repetitive units, may have been generated by assembling primordial exons together with their flanking regions during exon shuffling (reviewed in ref. 29). The close resemblance between the structural organizations of the preproenkephalin gene and the ACTH-β-LPH precursor gene suggests that these multi-hormone precursor genes may have evolved by a similar mechanism.

This investigation was supported in part by research grants from the Ministry of Education, Science and Culture of Japan, the Mitsubishi Foundation and the Japanese Foundation of Metabolism and Diseases.

Note added in proof: The cDNA sequence for human preproenkephalin reported recently by Comb *et al.*[30] confirms our assignment of the two introns, but indicates that nucleotides 71–157 (numbered beginning with the capping site) can be eliminated by RNA splicing. This spliced mRNA could not be detected by the probe or primer described in Fig. 3 legend, but we have now identified it as a major mRNA species in addition to the mRNA containing nucleotides 71–157 as a minor species. The 5′-terminal 113 nucleotides reported by Comb *et al.*[30] can be accounted for by an artefactual inverted repeat of the cDNA sequence starting with the second nucleotide of codon 242.

Received 2 March; accepted 2 April 1982.

1. Noda, M. *et al. Nature* **295,** 202–206 (1982).
2. Gubler, U., Seeburg, P., Hoffman, B. J., Gage, L. P. & Udenfriend, S. *Nature* **295,** 206–208 (1982).
3. Hughes, J. *et al. Nature* **258,** 577–579 (1975).
4. Stern, A. S. *et al. Proc. natn. Acad. Sci. U.S.A.* **76,** 6680–6683 (1979).
5. Kilpatrick, D. L., Jones, B. N., Kojima, K. & Udenfriend, S. *Biochem. biophys. Res. Commun.* **103,** 698–705 (1981).
6. Blobel, G. & Dobberstein, B. *J. Cell Biol.* **67,** 852–862 (1975).
7. Nakanishi, S. *et al. Nature* **287,** 752–755 (1980).
8. Nakanishi, S. *et al. Eur. J. Biochem.* **115,** 429–438 (1981).
9. Takahashi, H., Teranishi, Y., Nakanishi, S. & Numa, S. *FEBS Lett.* **135,** 97–102 (1981).
10. Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G. & Maniatis, T. *Cell* **15,** 1157–1174 (1978).
11. Benton, W. D. & Davis, R. W. *Science* **196,** 180–182 (1977).
12. Southern, E. M. *J. molec. Biol.* **98,** 503–517 (1975).
13. Lacy, E., Hardison, R. C., Quon, D. & Maniatis, T. *Cell* **18,** 1273–1283 (1979).
14. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65,** 499–560 (1980).
15. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. & Chambon, P. *Proc. natn. Acad. Sci. U.S.A.* **75,** 4853–4857 (1978).
16. Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. & Steitz, J. A. *Nature* **283,** 220–224 (1980).
17. Rogers, J. & Wall, R. *Proc. natn. Acad. Sci. U.S.A.* **77,** 1877–1879 (1980).
18. Berk, A. J. & Sharp, P. A. *Cell* **12,** 721–732 (1977).
19. Sollner-Webb, B. & Reeder, R. H. *Cell* **18,** 485–499 (1979).
20. Weaver, R. F. & Weissmann, C. *Nucleic Acids Res.* **7,** 1175–1193 (1979).
21. Breathnach, R. & Chambon, P. *A. Rev. Biochem.* **50,** 349–383 (1981).
22. Goldberg, M. thesis, Stanford Univ. (1979).
23. Benoist, C., O'Hare, K., Breathnach, R. & Chambon, P. *Nucleic Acids Res.* **8,** 127–142 (1980).
24. Sakano, H. *et al. Nature* **277,** 627–633 (1979).
25. Nakanishi, S. *et al. Nature* **278,** 423–427 (1979).
26. Numa, S. & Nakanishi, S. *Trends biochem. Sci.* **6,** 274–277 (1981).
27. Perler, F. *et al. Cell* **20,** 555–566 (1980).
28. Bell, G. I. *et al. Nature* **284,** 26–32 (1980).
29. Crick, F. *Science* **204,** 264–271 (1979).
30. Comb, M., Seeburg, P. H., Adelman, J., Eiden, L. & Herbert, E. *Nature* **295,** 663–666 (1982).
31. Weinstock, R., Sweet, R., Weiss, M., Cedar, H. & Axel, R. *Proc. natn. Acad. Sci. U.S.A.* **75,** 1299–1303 (1978).
32. Taii, S., Nakanishi, S. & Numa, S. *Eur. J. Biochem.* **93,** 205–212 (1979).
33. Rubin, C. M. & Schmid, C. W. *Nucleic Acids Res.* **8,** 4613–4619 (1980).
34. Houghton, M. *et al. Nucleic Acids Res.* **8,** 2885–2894 (1980).

A chicken histone H3 gene contains intervening sequences

James Douglas Engel & Barry J. Sugarman

Department of Biochemistry, Molecular and Cell Biology, Northwestern University, Evanston, Illinois 60201, USA

Jerry B. Dodgson

Departments of Microbiology and Public Health and Biochemistry, Michigan State University, East Lansing, Michigan 48824, USA

The organization of histone gene complexes in higher eukaryotes has long been a focus of study in eukaryotic molecular biology. The best studied and most extensively characterized histone gene families are those of the sea urchin *Strongylocentrotus purpuratus* and *Lytechinus pictus*[1,2], which are also the archetypal example of tandemly duplicated genes, each of the four core histone genes and the histone H1 gene being represented once in each tandem repeat unit. Recently, it has become apparent that the organization of histone gene families from higher eukaryotes does not follow these classical examples, a clustered but not tandemly duplicated arrangement having been found in chicken, mouse and humans[3–5]. In the chicken, the genes are organized in apparently random arrays with no cluster of genes being structurally analogous to another cluster[3]. In *Xenopus*, there is evidence for some degree of tandem duplication, and yet there is divergence within these clusters[6]. In yeast, the two gene clusters, each containing an H2a and H2b gene, are separate from other histone genes in the organism; furthermore these two clusters are unlinked to one another, but show evidence of close conservation of sequence between the two clusters[7]. During the routine screening and processing of a number of chicken histone gene recombinants we discovered one chicken histone H3 gene which displayed abnormal characteristics with respect to the majority of the genes in our collection. Closer examination of this gene showed it to be structurally different in that it is the first demonstrated example of a histone gene which contains intervening DNA sequences.

Two recombinants initially identified by hybridization to heterologous sea urchin and *Drosophila* histone gene coding sequences are shown in Fig. 1. Figure 1*a* shows the recombinant insert chromosomal DNA segment of λCH3d and Fig. 1*b* shows the position and orientation of an H3 gene contained within λCH3d and the subclone (pCH3dRI) used for further analysis. Figure 1*c* shows the sequencing strategy and an abbreviated restriction map in and around the H3 gene contained within λCH3d, and Fig. 1*d–f* shows the restriction map, subclone designation and sequencing strategy of the histone H3 gene in recombinant λCH4a. Mapping and hybridization details of λCH3d were published previously[3].

The histone H3 gene in λCH3d was chosen for sequence analysis because of strong suggestive evidence from RNA blots that this H3 gene was expressed in embryos (see below). In order to obtain a representative copy of each of the core histone genes for further identification of other, homologous chick histone genes within our isolated recombinants, we chose the H3 gene in λCH3d at random as an H3 gene-specific sequence. The H3 gene in λCH4a, however, was chosen for further analysis primarily because of strong hybridization of a mapping blot of this recombinant to cDNA prepared against adult red cell poly(A)$^+$ mRNA (data not shown). Since the only chick histone mRNA reported to be polyadenylated is histone H5 (ref. 8), we made the initial assumption that λCH4a contained a red cell-specific H5 histone gene. Subsequent sequence analysis (see below) proved that this particular locus encodes, rather, a novel histone H3 gene.

0028-0836/82/220434—03$01.00

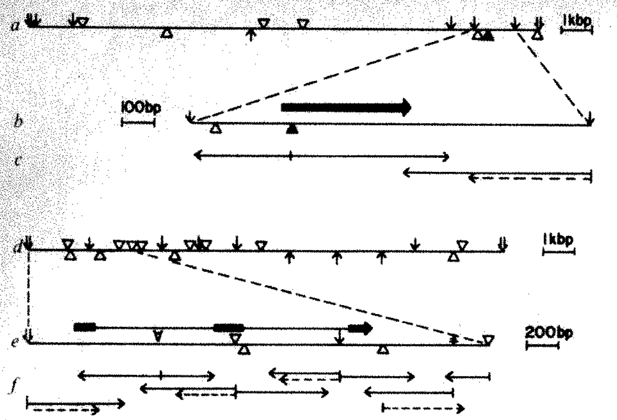


Fig. 1 Chick histone recombinants and DNA sequencing strategy. λ Charon 4A/genomic chicken DNA recombinants[19] λCH3d (ref. 3) and λCH4a are depicted in *a* and *d* respectively. Subclones containing the H3 genes from these chromosomal recombinants were prepared by ligation to homologous restriction sites of pBR322. The H3 gene in λCH3d is contained within subclone pCH3dR1 (*b*; heavy arrow indicates direction and extent of transcription), whereas the coding sequences of the H3 gene in λCH4a is contained within subclones pCH4a-2.0 (5′ and central exons) and pCH4a-2.3 (3′ exon). Restriction enzymes used were: ↓ = *Eco*RI; ↑ = *Bam*HI; ▽ = *Hin*dIII; △ = *Kpn*I; ▲ = *Sal*I; ‡ = *Hin*fI; ∀ = *Ava*I. DNA fragments used for sequence analysis were isolated by electroelution[20] and labelled at 5′ ends[21] (solid lines, *c* and *f*) or 3′ ends[22] (dashed lines, *c* and *f*). Chemical reactions[23] and polyacrylamide gel electrophoreses were performed as previously described[23–25].

As described previously[3], the H3 histone gene in λCH3d is linked to at least one other chicken histone gene, an H4 gene located in the 5′ direction (in transcriptional sense) from the H3 gene of λCH3d. However, no chicken histone genes other than the H3 gene shown in Fig. 1*e* have been identified on λCH4a (Fig. 1*d*) or on the chicken DNA sequences of an overlapping recombinant, λCH5d, which extends approximately 4 kilobases (5′ relative to the H3 gene) beyond the end of the λCH4a insert. However, we cannot yet rule out the possibility that an H1 or H5 histone gene is linked to this H3 gene, or that other histone genes are located in chicken chromosomal DNA close to this gene but beyond the λCH4a and λCH5d inserts.

The DNA sequence of the H3 genes in λCH3d and λCH4a is presented in Fig. 2. The H3 gene in λCH3d, like all other histone genes described to date, contains no introns (that is, the DNA sequence indicates that the polypeptide chain is colinear with the H3 polypeptide; Fig. 2, line B). This histone gene predicts an amino acid sequence which is identical to that of the adult chicken erythrocyte histone H3 polypeptide (line *a*) reported previously[9]. Furthermore, the gene contains 'CCAAT' and 'ATA' polymerase II consensus sequences 5′ to the gene (data not shown).

In striking contrast, the histone H3 gene represented by the chromosomal region expanded in Fig. 1*e*, is interrupted by two introns, dividing the coding sequence for this gene into three roughly equal coding parts. This DNA sequence (Fig. 2, line *c*) predicts an amino acid sequence very similar to that of a minor histone H3 polypeptide found in low abundance in somatic avian tissue (for example, in the erythrocyte, liver and perhaps other tissues)[10]. However, the predicted amino acid sequence of this gene differs by reciprocal, conservative changes from the chick erythrocyte histone H3.3 polypeptide previously described[11] at coding positions 31 and 87 (Fig. 2, line *d*). Thus we cannot be sure whether, in fact, this gene codes for the polypeptide initially described by Urban *et al.*[11], or perhaps codes for yet another (low abundance) erythrocyte histone H3 gene.

The DNA sequence of the interrupted gene contains no canonical sequences (outside of the coding sequence) indicative of an expressed gene for poly(A)$^+$ mRNAs isolated and characterized to date; that is, within 741 nucleotides 3′ to the termination codon we have been unable to find the 'AAUAAA' poly(A) addition sequence found in other eukaryotic poly(A)$^+$ mRNAs[12], nor within the 873 nucleotides 5′ to the initiation codon do we see evidence for 'CCAAT' or 'ATA' consensus sequences[13,14] (data not shown). However, the splice junction nucleotides 5′-GT and 3′-AG are present flanking both introns, as have been found in all intron-containing polymerase II-transcribed genes studied to date[15,16]. 255 bases 3′ to the termination codon, one finds an oligo d($A_{11}GA_9$) encoded in the mRNA sense strand. This oligo(A) sequence, if transcribed, would account for the selection of the mRNA synthesized from the split gene on oligo d(T) cellulose. We have no independent data to show whether, in addition, a post-transcriptionally added 3′ polyadenylate tail also exists.

To ascertain if (and if so, when) the H3 genes were expressed during avian development, RNA blotting experiments were undertaken. The results of this kind of experiment are shown in Fig. 3. The probe was generated from the histone H3 gene subclone of λCH3d used for sequencing (pCH3dR1; Fig. 1*b*) and hybridizes specifically and with high affinity to RNA prepared from several samples of 4 and 5 day embryonic RNA. Thus, this gene appears to be expressed both in the avian embryo and in embryonic red cells (Fig. 3, lanes 1, 2), but not in late embryonic or mature adult tissues (Fig. 3, lanes 3, 4). When the identical RNA samples were hybridized with similar probes which contained the split gene exons, hybridization was seen to all RNA samples examined (data not shown); however, the pattern of hybridization was more diffuse and always less

```
                                       10                                      20                                      30
a     AlaArgThrLysGlnThrAlaArgLysSerThrGlyGlyLysAlaProArgLysGlnLeuAlaThrLysAlaAlaArgLysSerAlaProAlaThrGly
b   atgGCGCGTACGAAGCAGACGGCGCGTAAGTCGACGGCGGGAAGGCGCCCCGCAAGCAGCTGGCCACCAAGGCGGCCCGTAAGAGCGCGCCGGCCACGGGC
c   atgGCCCGTACCAAGCAGACCGCCCGCAAGTCCACCGGCGGCAAGGCGCCCCGCAAGCAGCTGGCCACCAAGGCGGCCCGCAAAAGCGCGCCCTCTACTGGC
d                                                                                                    Ser

                  40                                      50                                      60
GlyValLysLysProHisArgTyrArgProGlyThrValAlaLeuArgGluIleArgArgTyrGlnLysSerThrGluLeuLeuIleArgLysLeuProPheGln
GGCGTGAAGAAGCCACACCGTTATCGGCCCGGTACGGTGGCGCTGCGCGAAATCCGGCGCTACCAGAAGTCCACGGAACTGCTGATCCGCAAACTGCCCTTCCAA
GGGGTGAAGAAGCCGCACCGCTACCGGCCTGGTACTGTGGCCCTCCGTGAAATCAGGCGCTATCAGAAGTCAACCGAACTTTTAATCCGCAAGCTTCCCTTCCAG
                         I1
                       766 bp
   70                                      80                                      90                                      100
ArgLeuValArgGluIleAlaGlnAspPheLysThrAspLeuArgPheGlnSerSerAlaValMetAlaLeuGlnGluAlaSerGluAlaTyrLeuValGlyLeu
CGCCTGGTGCGCGAGATCGCGNNNGACTTCAAGACTGATCTGCGCTTCCAGAGCTCGGCCGTCATGGCGCTCCAGGAGGCGAGCGAGGCCTACCTTGTGGGCCTC
CGTCTGGTGCGTGAAATTGCTCAGGACTTCAAAACAGATCTGCGCTTCCAGAGCGCTGCTATCGGTGCTTTGCAGGAGGCAAGTGAAGCCTACTTGGTTGGCCTG
                                                      Ala   IleGly       I2
                                                                       672 bp
                110                                     120                                     130       135
PheGluAspThrAsnLeuCysAlaIleHisAlaLysArgValThrIleMetProLysAspIleGlnLeuAlaArgArgIleArgGlyGluArgAla
TTCGAGGACACTAATCTCTGTGCCATCCACGCCAAGCGCGTCACCATTATGCCCAAGGATATCCAGCTCGCTCGCCGTATCCGTGGAGAGCGCGCCtga
TTTGAAGATACCAACCTGTGTGCTATCCATGCCAAACGTGTCACAATCATGCCAAAAGATATCCAGCTAGCACGCCGCATTCGTGGGGAGCGTGCCtaa
```

Fig. 2 DNA Sequence of linear and interrupted chicken histone H3 genes. The amino acid sequence of the major chicken erythrocyte H3 polypeptide (called H3.2 in ref. 11) is shown in *a*[9], and is also the sequence predicted by the H3 gene in pCH3dR1 (*b*). Unassigned glutamate/glutamine or aspartate/asparagine residues in the original polypeptide sequence[9] (between amino acids 73 and 85) are assigned from the DNA sequence (underlined third letter in three letter amino acid code; line *a*). The coding sequence of the split gene in λCH4a is shown in line *c* above the amino acid sequence for the minor adult H3.3 histone polypeptide[11] (identities in the two polypeptides (lines *a* and *d*) are deleted in line *d*). Differences in the two amino acid sequences are underlined; the differences between the prediction from the interrupted gene sequence (line *c*) and the H3.3 polypeptide variant[11] are boxed (positions 31 and 87). The position and size of the two introns in the split gene are shown as triangles.

intense than that of the uninterrupted H3 gene. Nonetheless, consistent hybridization was seen to the split H3 gene locus in RNA from every tissue which was examined. We conclude from these data that the histone H3 gene in λCH3d encodes an RNA species which is primarily embryonic in origin.

Thus the two chick histone H3 genes described here differ strikingly in primary DNA sequence. The H3 gene in λCH3d has a structure similar to other histone genes reported to date[2], and hybridizes exclusively to chick mRNAs which are embryonic in origin, whereas the H3 gene in λCH4a has two interruptions in the coding segment of the gene.

We have not been able to demonstrate the presence of either 5′ or 3′ consensus sequences common to other eukaryotic polymerase II transcripts in the DNA sequence of the split H3 gene. We are now examining sequences distal to the coding sequence portion of the gene reported here for the presence of these polymerase II consensus sequences. The absence of an observable poly(A) addition sequence (AAUAAA) is not surprising since this gene may not be polyadenylated by a post-transcriptional mechanism, and may appear in the poly(A)$^+$ mRNA fraction only because it contains a transcribed A-rich sequence ($A_{11}GA_9$). The absence of an 'ATA'-type 5′-flanking sequence is in contrast with several other chicken histone gene sequences examined (our unpublished results), but the actual mRNA initiation site may be outside the presently sequenced region of DNA. In particular, it is possible that another intron exists in the 5′ transcribed spacer region of the gene, and for this reason we have not yet identified the 'leader' sequence of the mature mRNA in the split gene.

It is interesting from a developmental viewpoint that the histone H3 gene in λCH3d encodes an embryonic message corresponding to a polypeptide which is likely to be identical to the major adult erythrocyte histone H3 (ref. 9), but nonetheless appears not to be expressed in the adult red cell. This suggests that a developmental switch occurs in histone biosynthesis in which an identical polypeptide is produced from two separate genes (one embryonic and one adult) which, from the RNA blots (Fig. 3) must differ significantly in nucleic acid sequence. The alternative explanation, that the two genes differ widely in amino acid sequence (for example, in the ambiguous glutamine/glutamate and asparagine/aspartate positions originally unassigned in the amino acid sequence[9]) is less compelling since the histones, as often reported, appear to show extreme amino acid sequence conservation[17].

The structure of the split histone H3 gene in λCH4a poses several important questions for further investigation. First, is this gene expressed? Several pieces of data suggest that the split gene is transcriptionally active: (1) cDNA prepared to poly(A)$^+$ chicken cellular RNA selected by oligo d(T) cellulose hybridizes exclusively and with high affinity to the presumptive exon sequences; (2) the DNA sequence predicts a polypeptide sequence which is very similar to that of an H3 histone found in adult chicken tissue; (3) no internal stop codons are found, in phase, in the three 'exons'; (4) the intervening DNA sequences are bounded by conventional (5′-GT, 3′-AG) splice junctions at the intron–exon borders; (5) in all tissues examined, RNA is transcribed which hybridizes to this locus. We are now examining the strandedness, tissue specificity, abundance and size of the transcription products at this locus. Our preliminary RNA blotting results raise the possibility that the pattern of chicken histone RNA synthesis might well be analogous to *Xenopus* 5S RNA gene expression, in which oocyte-specific 5S RNAs are synthesized only at early times (corresponding to our embryonic H3 gene in λCH3d), whereas the somatic 5S genes are expressed in both the oocyte and mature frog[18] (corresponding to expression of the split gene).

We thank Drew Rusling for discussion and help with computer analysis and Sheryl Niemiec, Patricia Creatura and Ed Diekhoff for technical assistance. We acknowledge access to the Stanford Molgen Project computer programs and data bank, made available through NIH resources and SUMEX, and helpful discussions thereof with L. Kedes. This work was supported by the NIH (J.D.E.: HL24415 and CA15145; J.B.D.: GM28837), the March of Dimes Birth Defects Foundation (J.B.D.: 1-0759) and the Leukemia Research Foundation; J.D.E. acknowledges the support of an American Cancer Society Faculty Research Award.

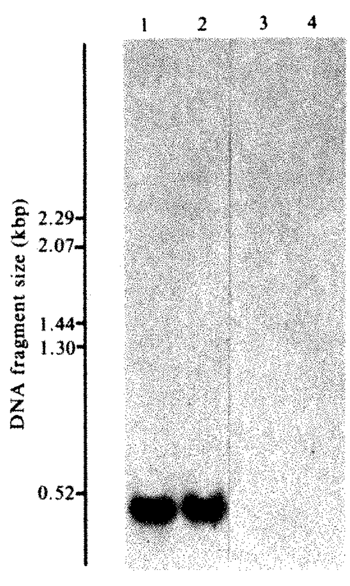


Fig. 3 Analysis of RNAs synthesized by the λCH3d histone H3 gene. Total cellular RNA and poly(A)$^+$ RNA was prepared as described elsewhere[26]; 5 μg per lane was reacted with glyoxal in dimethyl sulphoxide, electrophoresed and blotted to nitrocellulose[27]. RNAs used were: lane 1, 4-day chick embryo total cellular RNA; lane 2, 5-day chick embryo red cell cytoplasmic RNA; lane 3, adult chicken red cell cytoplasmic RNA; lane 4, oligo d(T)-cellulose selected adult chicken red cell cytoplasmic RNA. Hybridization was to nick-translated pCH3dRI (specific activity: 1.5×10^8 c.p.m. per μg). After hybridization, filters were finally washed in a solution containing total monovalent cation concentration of 20 mM. Thus the homologous T_m should be about 77 °C for a 58% G+C base-pair hybrid[28]; the filter washes were performed at 60 °C. Exposure time was 9 h.

Received 28 December 1981; accepted 21 April 1982.

1. Kedes, L. H. *Cell* **12**, 321–331 (1978).
2. Hentschel, C. C. & Birnsteil, M. L. *Cell* **25**, 301–313 (1981).
3. Engel, J. D. & Dodgson, J. B. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2856–2860 (1981).
4. Sittman, D. B. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 4078–4082 (1981).
5. Heintz, N., Zernik, M. & Roeder, R. G. *Cell* **24**, 661–668 (1981).
6. Zernick, M., Heintz, N., Boime, I. & Roeder, R. G. *Cell* **22**, 807–815 (1980).
7. Hereford, L., Fahrner, K., Woolford, J., Rosbash, M. & Kaback, D. B. *Cell* **18**, 1261–1271 (1979).
8. Molgaard, H. V., Perucho, M. & Ruiz-Carrillo, A. *Nature* **283**, 502–504 (1980).
9. Brandt, W. F. & von Holt, C. *Eur. J. Biochem.* **46**, 419–429 (1972).
10. Urban, M., Franklin, S., Goldman, P. & Zweidler, A. *J. Cell Biol.* **79**, 119a (1978).
11. Urban, M. K., Franklin, S. G. & Zweidler, A. *Biochemistry* **18**, 3952–3960 (1979).
12. Proudfoot, N. J. & Brownlee, C. C. *Nature* **263**, 211–216 (1976).
13. Efstradiatis, A. *et al. Cell* **21**, 653–668 (1980).
14. Goldberg, M. L. thesis, Stanford Univ. (1979).
15. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. & Chambon, P. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4853–4587 (1978).
16. Marx, J. L. *Science* **212**, 653 (1981).
17. Isenberg, I. *A. Rev. Biochem.* **48**, 159–191 (1979).
18. Ford, P. J. & Brown, R. D. *Cell* **8**, 485–493 (1976).
19. Dodgson, J. B., Strommer, J. S. & Engel, J. D. *Cell* **17**, 879–887 (1979).
20. Girvitz, S. C., Bacchetti, S., Rainbow, A. J. & Graham, F. L. *Analyt. Biochem.* **10**, 492–496 (1980).
21. Maxam, A. & Gilbert, W. *Meth. Enzym.* **65**, 499–560 (1980).
22. Challberg, M. D. & Englund, P. T. *Meth. Enzym.* **65**, 39–43 (1980).
23. Smith, D. R. & Calvo, J. M. *Nucleic Acids Res.* **8**, 2255–2273 (1980).
24. Dolan, M., Sugarman, B. J., Dodgson, J. B. & Engel, J. D. *Cell* **24**, 669–677 (1981).
25. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
26. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. *Biochemistry* **18**, 5294–5299 (1979).
27. Thomas, P. S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201–5205 (1980).
28. Schildkraut, C. & Lifson, S. *Biopolymers* **3**, 195–210 (1965).

BOOK REVIEWS

Nuclear power and the people

Eric Ashby

IN January 1977 the Secretary of State for Energy issued a paper entitled *Some Aspects of the Safety of Nuclear Installations in Great Britain*. The importance of this paper lies not in its content but in the reasons for writing it; for it was in part a response to forty-five questions put by the Friends of the Earth. In John Chicken's words, it "established the precedent for questions by interest groups to the Government on nuclear safety to be answered publicly".

Chicken's book recounts the brief history of public concern about hazards from nuclear power and the effects of this concern upon policy. Do not be put off by the elephantine adjectival nouns in the title: the book is a clearly written for the general reader and well documented for the specialist who may want to consult the sources. Its theme is that the terrors of nuclear warfare have, by association, coloured attitudes towards the real, but quite different, risks in the nuclear power industry. The effect of this has been a preoccupation with safety forced upon the Establishment, impulsive reaction — fanned by the media — to every report of an accident and a refusal to accept the evidence that, so far, it has been safer to generate power from uranium than from coal or oil. The climax of the book is an analysis of the Windscale Inquiry published in 1978 and the value of the book is that it puts this Inquiry into historical perspective. It explains (what might otherwise be inexplicable) how nuclear power stations are regarded by some people as polluting in the ancient meaning of the word "pollute": "to render morally impure, desecrate, destroy the sanctity of . . .". As was demonstrated at the Inquiry, the very use of radioactive materials for peaceful purposes is tainted by their possible use for deliberate destruction.

Chicken's sober and modest chronicle begins with a summary of the development of nuclear power and the ways in which it is now controlled. Against this background he sets his story. First, there was the development of a British atomic bomb, described in the superb official history by Margaret Gowing. The decision to make the bomb was taken by a Labour Government under Attlee; or, rather, by a small group of men working in such secrecy that (it is said) even the Cabinet was "excluded from discussion of many of the major decisions on atomic policy". During the whole of the Labour Government's period of office (writes Chicken) "from 1946 to 1951 there was not a single debate

Nuclear Power Hazard Control Policy. By John C. Chicken. Pp.272. Hbk ISBN 0-08-023254-X; pbk ISBN 0-08-023255-8. (Pergamon: 1982.) Hbk £20, $40; pbk £7.50, $15.

on Atomic Energy in the House of Commons . . .".

It seemed at first as though the peaceful uses of atomic energy could be developed with as little participation from the public as had occurred in its use for defence. The Atomic Energy Authority was set up in 1954. Its first report evidently did not anticipate hostility from the public. It carried a mild reassurance about risk: the reactors would "present no more danger to people living nearby than many existing industrial works that are sited within built-up areas". In 1957 Britain's first nuclear power station was opened and it was announced that more would follow; a Safety Branch of the Authority was created: all without any serious signs of concern among the public. It was a period of "quiet acceptance"; policy for energy was being left to the experts.

In the 1960s this quiet acceptance was overturned by a worldwide surge of disenchantment with technocrats and the establishments that employed them. The episode is too close to us for an objective historical analysis to be made of it. Chicken concentrates upon its impact on policies for energy, though he quite rightly sees that nuclear power was only one of several issues used by the so-called "counter culture" as tools for protest. The movement began in the United States before it reached Britain, fuelled by resentment about the Vietnam war and the denial of civil rights to minority groups. It spread to an indictment of industry, symbolized by Ralph Nader's withering criticism of negligence over safety in the design of automobiles. A wave of anxiety about pollution and a mood of guilt about exploitation of the environment swept across the affluent industrial nations. Radiation is the most ominous form of pollution; its military uses, its temptation for terrorists, its association with cancer, its persistence if released into the environment, the delayed action of its alarming toxicity, even its invisibility: all combined to make nuclear power a prime target for suspicion and protest.

Anti-nuclear lobbies work at two quite different levels, often simultaneously. At one level they question the need for nuclear power at all; they want an energy policy based on coal or, better, renewable resources, and much greater economy in the use of energy. They are willing to accept a lower material standard of living to achieve these ends. At the other level are those who question the safety of specific reactors on specific sites. It is at this second level that assurances about hazard-control become essential if there is to be a viable nuclear power industry. In 1975 two events occurred that concentrated attention upon hazard-control: one was the accident to a reactor at Browns Ferry in Alabama which — though nobody was hurt — was useful ammunition for the anti-nuclear lobbies. The other event was the publication of the Rasmussen Report, which became — and remains — the most useful weapon in the armoury of pro-nuclear lobbies. At the time of the Presidential Election in the United States in 1976 the call for a moratorium in the building of nuclear reactors became an election issue. In the following year tens of thousands of demonstrators challenged the building of a fast reactor in France. In Britain there was comparative quiet on these issues until British Nuclear Fuels proposed to extend their fuel reprocessing plant at Windscale. The Government wisely decided to throw the whole issue open to public scrutiny and discussion.

There is already a shelf-load of books, articles and pamphlets about the Windscale Inquiry. Nevertheless Chicken's careful analysis of the decisions about risk and the influence of public interest groups on these decisions is a valuable addition to this literature. The important and encouraging point he makes is that the public interest groups, despite their reliance on voluntary subscriptions to pay the costs of conducting their case, did have a significant influence on the findings of the Inquiry and taught the proponents of the proposal lessons they would not have learnt without the dedicated and competent efforts of such bodies as the Friends of the Earth.

Chicken's analysis leads him to make some tentative suggestions about procedures for involving the public in decisions about what they regard as "acceptable risks". Despite criticisms (some of which, for example in *The Ecologist*, quoted on p. 172, were contemptible) the style of the Inquiry was satisfactory. But there is room for improvement in three respects. First, it is important that public interest groups should be financed from public funds for the significant part they play in such inquiries. Chicken does not mention the

successful way in which this has been done for similar inquiries in Canada. Second, it is essential that the public interest groups should have access to all the technical data possessed by the proponents. Third, and equally essential, Parliament, as the legally constituted forum for the ventilation of public opinion, should participate in decisions as to how safe is safe enough.

Finally, a word about the style and production of the book. John Chicken is to be congratulated on the impeccable way in which he refers the reader back and forth to the details of his complex chronicle: the book is admirably signposted. As to production, the book has been published from camera-ready copy which is a good thing because it keeps the price at a reasonable level; but there are a lot of misprints: surely these could have been corrected as easily in typescript as in ordinary set type. □

Eric Ashby is a Fellow of Clare College, Cambridge. He is joint author with Mary Anderson of The Politics of Clean Air *published by the Clarendon Press late last year.*

Journals' review issue 1982

On October 7th *Nature* will publish a second review supplement devoted to science journals. Last year's supplement, covering journals first published between January 1978 and May 1980, appeared in *Nature* **293**, 341-369; see p.343 of that issue for details.

Criteria for inclusion of a journal in the 1982 issue are that:

- the first number appeared, or the journal was re-titled, between June 1980 and May 1981;
- it is published at least three times a year;
- the main language used is English.

Broadly, periodicals of professional interest to scientists will be considered for review, with the exception of abstracts' journals.

In addition it is hoped to cover publicly available newsletters, first published between January 1978 and May 1981, in that issue.

Publishers and societies are invited to submit four sample issues of periodicals satisfying the above criteria, *including the first and most recent numbers*, to the Review Editor, *Nature*, 4 Little Essex St, London WC2R 3LF, England (London 836 6633 ext 2570) as soon as possible.

● **Oliphant** by Stewart Cockburn and David Ellyard (reviewed in *Nature* **296**, 685; 1982) is available in the UK from W. Heffer, 20 Trinity St, Cambridge CB2 3NG or C.M. Watkins, 44 Madrid Rd, London SW13. Price is £14.90 + post and packaging.

●Vernon Booth's **Writing a Scientific Paper and Speaking at Scientific Meetings, 5th Edn** (reviewed in *Nature* **296**, 499; 1982) can be obtained from The Biochemical Society Book Depot, PO Box 32, Commerce Way, Colchester CO2 8HP, UK.

Catastrophes in science and engineering

Colin Upstill

Catastrophe Theory for Scientists and Engineers. By R. Gilmore. Pp.680. ISBN 0-471-05064-4. (Wiley: 1981.) £32.15, $61.
Instabilities and Catastrophes in Science and Engineering. By J.M.T. Thompson. Pp.226. Hbk ISBN 0-471-09973-2; pbk ISBN 0-471-10071-4. (Wiley: 1982.) Hbk £14.50, $34.75; pbk £8.25, $19.50.

THE term "catastrophe theory" is open to wide interpretation, depending upon the scope accorded it. René Thom considered catastrophe theory to be a theory of general morphology, a topological bifurcation theory for systems and bifurcations of very general type, and this usage is now common. Under this broad roof lives a great body of work including such topics as bifurcation theory, dynamical systems theory, singularity theory and elementary catastrophe theory. The lines of demarcation between topics are neither fixed nor of zero width: something of a consensus is emerging, but many authors have used, do use, and no doubt will use, these and other terms without clearly stating what they mean by them.

Elementary catastrophe theory (ECT), the study of the equilibria of gradient systems, is the part of catastrophe theory which embodies the most simplifying assumptions, and is the most developed and accessible part of the theory. The qualifying adjective is often dropped for ease of exposition; thus catastrophe theory may mean something very general or something quite specific. When, as sometimes happens, the same author uses the term in both senses, the potential for confusion is maximized. Neither Gilmore nor Thompson is actually guilty of this, but there is a marked difference in the manner in which they deal with the welter of jargon.

In the first three pages of an excellent text, Gilmore explains briefly and clearly what catastrophe theory is, and what dynamical and gradient systems are and how they and ECT fit together. In the first of four parts of the text, he goes on to "present Elementary Catastrophe Theory from an elementary point of view" in a way which is well tailored to the needs and background of most physical scientists and engineers, defining technical terms crisply and clearly without recourse to the cryptic notations and abstraction of the mathematical literature.

Thompson's book is not so much a text as a review of many (but by no means all) current applications of catastrophe theory in the physical sciences and in engineering. It performs this task splendidly, provided that one has some knowledge of the subject to start with. For, despite the claim of the preface that the presentation is "at the level of popular magazines such as *Scientific American* or *New Scientist*", a great variety of jargon, both catastrophe theoretic and specific to the specialist fields covered, is used without ever being explicitly defined. The first use of many such terms is flagged by their being italicized, but setting something in italics is no substitute for explaining what it means. In collaboration with G.W. Hunt, Thompson is preparing a companion book to deal with much of the underlying theory, so it may well be that the decision to use unexplained technical terms is a deliberate one. Nonetheless, I believe that the book is considerably less accessible to many of the intended readers than it would have been with the occasional extra sentence of definition and explanation. This problem is exacerbated by Thompson's use of a transcript of his inaugural lecture as a basis for the introductory chapter, which gives a brief historical survey of the subject and an overview of the contents of the rest of the book. This is excellent for the *cognoscente*, but at the price of presenting the reader new to the subject with a high density of unexplained technical terms, and with a number of complicated diagrams referred to by phrases such as "the familiar such-and-such", which will almost certainly be anything but familiar to him.

Having taken Thompson to task for not writing the book the cover and the preface lead one to expect, let me repeat that the book he has actually written is a good one, well illustrated — essential when dealing with so many intrinsically geometric notions — and with copious references. He distinguishes clearly between static and dynamic bifurcations whilst showing how they are related, particularly in the context of engineering structures: among the other instability phenomena reviewed are gravitational stellar collapse, atomic lattices, biochemical reactions (including the celebrated Brusselator), population dynamics and ecology, hydrodynamical turbulence and the chaotic dynamics of strange attractors.

In Part 2 of his text, Gilmore considers some applications of ECT, and in Part 3 he first takes us beyond ECT into dynamical systems theory and then considers some examples. These parts of his text are, broadly speaking, complementary to Thompson's book; there is some overlap, but most of the topics Gilmore discusses are not considered at any length by Thompson, and vice versa. Gilmore deals with such topics as phase transitions in classical and quantum mechanical systems, climatology, and optical caustics and diffraction patterns. His discussion of optics is noticeably out of date in comparison with that of other topics, but he neither makes nor implies any claim to be reviewing the state of the art, giving few references and more relaxed and yet deeper discussions than Thompson.

Indeed, Gilmore's relaxed style will not be to everyone's taste; for example, after a page or two describing a standard treatment of some phenomenon, he concludes with the acerbic aside "In short, an intractable problem is replaced by an intractable problem". It is a style he handles well; it could so easily be overdone, but in my view it is not, and greatly enhances readability and thence understanding. The most formal section, also the shortest, is Part 4, in which he presents in some detail the mathematics underlying ECT. In contrast with Part 1, this is (necessarily) couched in sophisticated mathematical language. It would have been easy to have included this material in Part 1, and many readers would not then get beyond it: it is greatly to the benefit of the book that this mistake was avoided. Coincidentally, Part 4 is the least satisfactory part of the book, when judged against the standard work in the field, the classic text of Poston and Stewart (*Catastrophe Theory and its Applications;* Pitman, 1978) whose treatment of Thom's theorem, transversality, determinacy and unfoldings stands head and shoulders above Gilmore's. That criticism is more than offset by the fact that, particularly for the reader not well versed in the language of modern pure mathematics, the first nine-tenths of Gilmore's text is excellent.

In his epilogue, Gilmore asks the question "Is catastrophe theory useful and important?" and concludes by asserting his belief that it is. I share that view, and both his and Thompson's books are convincing evidence in support of it. □

Colin Upstill is at the H.H. Wills Physics Laboratory of the University of Bristol.

All around andesites, and beyond

Alexander R. McBirney

Orogenic Andesites and Plate Tectonics. By J.B. Gill. Pp.390. ISBN 3-540-10666-9. (Springer-Verlag: 1981.) DM 98, $44.60.

A THOROUGH, up-to-date review of orogenic andesites has long been needed and is especially welcome at this time, when many petrologists are taking stock of this group of rocks and their interpretation in terms of accepted concepts of plate tectonics. In undertaking this formidable task, James Gill has performed a notable service to petrology. His book is certainly the most comprehensive compilation of data and ideas that has yet been attempted.

The book summarizes almost every aspect of the occurrence, composition and genetic relations of andesites. It reviews their structural and tectonic setting, rheological properties, eruptive behaviour, geochemical and mineralogical features, compositional variations in time and space, and all the major petrogenetic theories proposed to explain their relations to plate tectonics. This mass of data is presented in a concise, well-organized fashion. In his evaluation of the various schemes which have been proposed to account for orogenic igneous rocks, Gill turns much of his attention to their relationship to subduction, and a large part of the book is an attempt to reconcile the occurrence and unusual features of the rocks with a party-line view of plate tectonics. In doing so, he has made a conscientious effort to mention all observed relationships, together with every known exception; the result is a text which tends towards the encyclopaedic, with awkward sentences that one must read two or three times to sort out an elusive trivial point. The style improves, however, when Gill comes to a question of interpretation that clearly interests him. The text then blossoms and becomes quite readable.

The book is burdened by the long-standing difficulty of defining what rocks should be included under the name andesite. The reader encounters this problem in the very first sentence, which tells him that "active volcanoes on Earth erupt andesite more than any other rock type". Before recovering from this, he reads, further down the same page, that these andesitic volcanoes include Kilimanjaro and Hekla! The explanation for these startling assertions soon becomes apparent when Gill gives his definition of andesite as any hypersthene-normative rock with a silica content between 53 and 65 weight per cent. Although this definition is offered in the interests of "simplicity" the resulting confusion in the mind of the reader persists throughout the book, for one is often in doubt as to whether, in referring to "andesites", Gill means calc-alkaline andesites or some other intermediate sub-alkaline rock. Much needless discussion is devoted to rocks that have no relevance to andesitic volcanism as most geologists perceive it.

There is a certain logic in including among andesites the important group of intermediate tholeiitic rocks of island arcs that are clearly related to subduction and share many genetic features in common with the "true" calc-alkaline andesites with which they are closely associated, sometimes even in the same volcano. It seems illogical not to refer to these rocks as tholeiitic andesites, but most petrologists would certainly agree that they are quite distinct from rocks of similar composition in settings such as Iceland, the Galapagos Islands or East Africa. The problem is made more difficult by the way in which Gill attempts to distinguish tholeiitic and calc-alkaline andesites. While he recognizes the importance of iron content, he ignores the more reliable compositional criteria of Irvine and Baragar and makes no mention of the high alumina and normative plagioclase contents that many petrologists consider compositionally diagnostic and crucial to any petrogenetic interpretation of calc-alkaline andesites.

Because much of the evidence bearing on these rocks comes from geochemical relations, any book of this kind must deal with complex trace-element patterns and isotopic ratios, and with how these may be related to the processes of magmatic evolution. Gill uses these data to weigh the relative importance of melting subducted oceanic lithosphere, assimilating crustal material and differentiating basaltic magma. In this way he systematically assesses essentially all the current ideas on the origin of andesitic magmas. While he does not hesitate to give his own judgement on these questions, he does so in the context of an overall appraisal of the evidence, so that readers can accept or reject Gill's conclusions according to their own view of the facts.

No two persons would agree as to how much weight should be given to various types of evidence bearing on the origins of these rocks. I would personally have preferred more data on geologic relations and less on trace-element geochemistry. It astonishes me that anyone can write an entire book about andesites and never present a single geologic map or structural section of an andesitic volcano. The book, like most modern work which covers orogenic igneous rocks, deals almost exclusively with the structural relations observed today around the Circum-Pacific and gives much less attention to the geologic past. If it is important to note that all active andesitic volcanoes are in well-defined belts close to convergent plate boundaries, is it not equally important to stress that the same has not always been true in the past? Many rules of thumb based on modern relations fail when we look beyond the Quaternary, and I suspect that one can learn more about andesites from the anomalous cases that violate our conventional models than from all the examples that neatly fit them.

These, of course, are petty differences of view, and they detract in no way from my unreserved recommendation of this book as a timely, scholarly survey of a very complex subject. The sincerest praise I can offer any book is to say, as I can of this one, that I have adopted it as a text in my graduate-level petrology course. No student of orogenic igneous rocks should neglect adding it to his professional library.

Alexander R. McBirney is a Professor in the Department of Geology at the University of Oregon, Eugene.

How ionizing radiation affects cells

R.B. Setlow

The Molecular Theory of Radiation Biology. By K.H. Chadwick and H.P. Leenhouts. Pp.377. ISBN 3-540-10297-3/0-387-10297-3. (Springer-Verlag: 1981.) DM 128, $67.20.

SOME of the best epidemiological data relating carcinogenic risks to environmental hazards comes from research into the consequences of exposure to ionizing radiation. The recent recalculation of dosimetry at Hiroshima and Nagasaki has just about eliminated any effects in these populations ascribable to neutrons, and, because gamma ray data are not robust below 30 rads, there are, for practical purposes, no human data at low doses or for particles of higher linear energy transfer. Thus, even in this best case, one needs models and theories for extrapolating from acute high dose effects to chronic low dose ones. If such a problem cannot be solved for ionizing radiation with its relatively good dosimetry, the chances are negligible that it will be solved for populations exposed to chemical agents.

In this book, Chadwick and Leenhouts describe the development of a theory that can be used to extrapolate the biological effects of ionizing radiation to low doses, many different LET particles, and to interactions between radiation and chemical agents for end points such as cytotoxicity, mutation, chromosome aberrations and carcinogenesis. Their theory is based on the premise that DNA is the target and that double-strand breaks in it are the molecular lesions reponsible for the biological effects of radiation. The authors are led to this assumption for a number of reasons, the principal one of which is that the dependence of many biological effects on dose contains both a linear and a square term. The linear term is thought of as representing double-strand breaks arising from the passage of a single ionizing particle close to DNA, and the square term represents double-strand breaks arising from two nearby single-strand breaks made by two ionizing particles.

The text is a careful and complete exposition of the implications of these assumptions. Of necessity, other concepts and hence additional parameters and suppositions must be introduced, such as allowance for the repair of double-strand breaks and the assumption that *all* radiation mutations arise from double-strand breaks. Thus, the analysis rapidly accumulates other variables that are often difficult to evaluate independently and that must be adjusted to fit the experimental results. The general problem faced by the authors is equivalent to that of determining enzyme mechanisms at the molecular level from enzyme kinetics: it cannot be done. Certainly, there is no unique solution to the problem especially because there have been relatively few good determinations of double-strand breaks as a function of dose and, because of a lack of basic knowledge of events at the molecular level, of how ionizing radiation affects even simple end points — such as the inhibition of the initiation of clusters of eukaryotic replicons — and inhibits the progression of cells through the cell cycle.

The authors do themselves an injustice by overstating their case. The assumption made early in the book that double-strand breaks are important becomes a fact later on. Such confidence is misplaced because it ignores data on the effects of radiation on the replicative form of ΦX174, in either protective or dilute solution, in which inactivation is not correlated with double-strand breaks but with some type of base damage. Moreover, although fibroblasts from individuals with the disease ataxia telangiectasia are more sensitive to the cytotoxic effects of ionizing radiation than are normal cells, and show more chromosomal aberrations as a result of radiation, they are hypomutable. Hence, not all of the biological end points are comparable and the simple theory expounded in the text seems to need major revision based on solid biology.

These are obviously serious criticisms. Nevertheless *The Theory of Radiation Biology* is an important book for radiation and environmental biologists for, if nothing else, it is sure to stimulate new experiments and ideas. □

R.B. Setlow is a Senior Biophysicist at the Brookhaven National Laboratory and a member of the Board on the Effects of Radiation of the US National Academy of Sciences.

Counting calories

John R. Krebs

Physiological Ecology: An Evolutionary Approach to Resource Use. Edited by Colin R. Townsend and Peter Calow. Pp.393. Hbk ISBN 0-632-00555-6; pbk ISBN 0-632-00617-X. (Blackwell Scientific/Sinauer: 1981.) Hbk £21, $38; pbk £11.80, $23.60.

I WAS once told by a fellow graduate student, trained in herpetology, that physiological ecology was the science of shoving thermometers up lizards'cloacas. A quick check in the subject index of this volume revealed that, given the veracity of my colleague's assertion, things have come a long way in the past 15 years.

The book, comprising a collection of original articles and intended as a text for undergraduate courses, is mainly about how plants and animals get their daily calories and what they then do with them. The unifying theme linking such diverse topics as photosynthesis, digestion, cellular repair and social behaviour is that all aspects of earning and spending calories can be treated from a functional viewpoint as optimization problems. Different chapters illustrate different degrees of progress in applying this approach. In some areas, for example leaf design and foraging behaviour, the first generation of explicit models has been formulated and tested fairly thoroughly, while in others such as defence mechanisms and social behaviour the questions have, in the main, been formulated only qualitatively.

The articles also vary in the extent to which they present totally new ideas. R.M. Sibly, in one of the more original chapters, provides a fresh and stimulating approach to digestive strategies. He starts with a version of the marginal value theorem to predict optimal retention times for food in the gut, and leads via a discussion of gut morphology to consider why food empties from the stomach at an exponentially decreasing rate from the time of the last meal (no simple answer to this question emerges, but the relevant data and hypotheses are neatly brought together). The highlight of the book in terms of entertainment value is D.H. Janzen's fireside chat about defences against predators and parasites. Among the aphorisms that abound in this chapter is the answer to "Why should some species of seed be so well defended when the plant makes so many?" — ". . . for the same reason that a young man protested at being sent to Vietnam". At the same time, I should add that Janzen does not, to my mind, succeed in pushing the study of plant and animal defences from a series of intriguing evolutionary stories to forming a coherent theory. He appeals for a change in attitude from "descriptions of yet another alkaloid [to measuring] fitness gains and costs that the organism accrues with and without its defences" without stating explicitly how these hard-won measurements will be used to make a general theory of defences.

Some contributors — for example Pianka on resource acquisition and allocation among animals, and Gosling and Petrie on the economics of social organization — try to cover too wide a field, and in so doing do not analyse any single problem beyond introductory textbook level, while others — Kirkwood on repair and its evolution, for example — deal in detail with a more restricted range of topics. Overall, however, the book is a worthy collection of reviews which covers a range of subjects that are not normally assembled in one volume. □

John R. Krebs is a Nuffield Foundation Science Fellow at the Edward Grey Institute of Field Ornithology and E.P. Abraham Fellow of Pembroke College, Oxford.

The principles of vision enumerated

Oliver Braddick

A Taxonomy of Visual Processes. By William R. Uttal. Pp.1,097. ISBN 0-89859-075-2. (Lawrence Erlbaum: 1981.) $100, £50.

THE stated aim of Uttal's book is to "... systematize the data base of perceptual psychology into a comprehensive intellectual scheme ...". When the route to this goal runs through 1,100 pages and over 3,000 references, it must involve to a fair degree the exposition of the "data base" itself, and there is here a generous view of its extent, including the discovery of alpha particles, how a cathode ray tube works, the embryology of frog rods and the theory of rainbows. More conventionally, the field of vision is covered from photometry to geometrical illusions. It would not be reasonable to expect a single author to compete in authority over this range with the several multi-author handbooks of similar scope that have appeared in recent years, and Uttal does not do so.

The author shows a propensity for organizing his data base in lists, which rather blunt the reader's appreciation of what might be the relative significance of the information. This culminates in the closing pages in a list of 67 "general principles" from which we may sample No.4:

> For millennia, philosophers, scientists, and laymen alike have felt a deep-seated urge to explain the phenomena of perception. . . . No perceptual theory, however, has ever withstood the test of time . . .

and No.59: "The magnitude of many illusions depends on the orientation of the pattern . . .". Several of the lists contain direct contradictions that remain unresolved and without comment by Uttal. The general air is one of abstracting rather than digesting the scientific literature; opportunities for integration are conspicuously missed, for example the use of the temporal modulation transfer function to express data on the visual response to a variety of time-varying stimuli.

But it would be unfair to dwell on Uttal's presentation of the data base, rather than his intention of providing a framework for systematizing it, the "taxonomy" of the title. This taxonomy is one of sequential levels of processing, from Level 0 in the optics of the environment and the eye, to Level 5 which consists of active, attentive manipulations of perceived information which are reserved for another projected book. Most visual scientists use some such idea of levels, albeit informally, as a way of compartmentalizing problems. This is not to say that it is the best principle for the organized presentation of our knowledge of visual perception, and several disadvantages become apparent in Uttal's attempt to do so. One is that there are many phenomena about which we know a good deal, and which are clearly important, but which we cannot partition among levels with any certainty. Ironically, Uttal's own experimental work largely falls into this category, and into a chapter on temporal interactions which he has to present as a frank digression, unassignable to any of his levels on present understanding. In other cases, the organization of the book by levels makes it difficult to evaluate alternative explanations critically. For example, the issue of lateral inhibition versus more global processes in brightness contrast is raised, with vigorous assertions of the importance of the latter, but the plan of the book does not allow them to confront each other. Another consequence is an absence of any functional perspective. The purposes for which vision is used cut right across these levels, so they get little attention here.

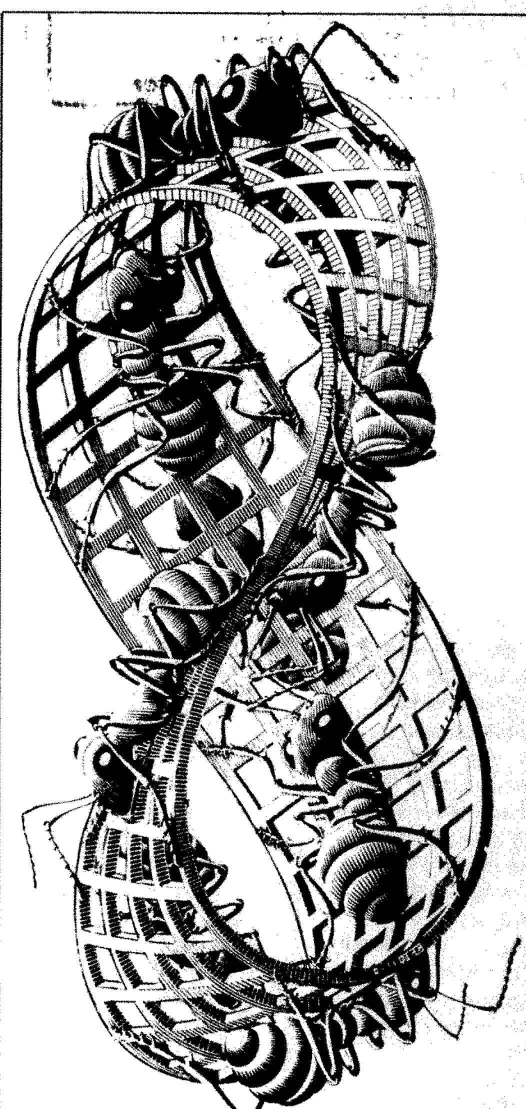
A reason for the taxonomic approach becomes apparent in a chapter entitled "Mezzolog: On the Limits of Neuroreductionism". Uttal is concerned that much current theorizing, providing speculative neural bases for perceptual phenomena, is premature. Once again, this view is expressed centrally as a list: "Questionable Dogmas", Nos 1–18. Some of the dogmas are general suppositions about the relation between brain activity and perception, and indeed some of these are quite widely believed, at least implicitly. Others (for example No. 15: "Metacontrast results from peripheral lateral inhibitory interactions that diminish the strength of the conducted neural signals") may be specific instances of neuroreductionism, but are not necessarily widely accepted even by those committed to neural reductionist explanations of the phenomenon concerned. Dogmas of these two kinds need rather different kinds of critique, but it is not clear that the difference is appreciated.

The role of the taxonomy is to draw a sharp line between Levels 1–2, where Uttal offers accounts in terms of interaction of neural signals, and Level 3–4 where the processes are "interpretive", the neural representations involved are not at all understood, and theories are psychological or purely formal.

Uttal's charge of premature reductionism is just in many cases. However the structure and style of his book do not favour a critical marshalling of the arguments on the neural basis of specific visual phenomena. Indeed, given the scale of the book, any particular issue is pursued in surprisingly little depth. The temptation of encyclopaedic breadth seems to have deflected the author's purpose of exploring the limits of current neural explanation, which is a pity, for that purpose is a very worthwhile one.

Uttal recounts in his preface that when he "chose to become an explicit generalist rather than a totally dedicated ... laboratory scientist" the reaction of some of his colleagues was such that "some of the experiences encountered were non-supportive to say the least". He defends his role as one of "analysis, criticism, integration and synthesis". I fear that generalism has caused these activities to be diluted and dispersed to a rather unsatisfying degree. □

Oliver Braddick is a Lecturer in Experimental Psychology at the University of Cambridge.



The art of M.C. Escher: this picture, Möbius Strip II (Red Ants), is reproduced from *Escher*, edited by J.L. Locher and recently published by Thames & Hudson, London, and Abrams, New York. The book, containing both a copiously illustrated catalogue of Escher's graphic work and a biography, costs £35, $65.

Inner machinery of the silicon chip

Andrew Holmes-Siedle

Instabilities in MOS Devices. By J.R. Davis. Pp.190. ISBN 0-677-05590-9. (Gordon & Breach: 1982.) $24.50.

MOST silicon integrated circuits are smaller in size than a fingernail and contain many thousands of working elements, the most common of which is the metal-oxide-semiconductor (MOS) transistor. In this, all of the important electronic activity takes place in a minute volume, a layered structure composed of crystalline silicon, thermally-grown silicon dioxide and a conductor. Attempts are always being made to reduce the unit size of the transistor element, but during such development it is crucial to maintain the performance of the intimately interwoven materials of the layered structure. *Instabilities in MOS Devices* is a discussion of the physical processes occurring in this metal-oxide-silicon system.

In MOS structures under electrical fields, electrons, holes and ions sometimes migrate. This movement produces the instabilities in electrical properties referred to in the title. In their constant endeavour to ensure that instability effects do not wreck new designs, physicists regard the silicon dioxide layer as the key element; it is a complex film but, in this work, can often be regarded as a simple, wide-band-gap insulator region, sandwiched between the "electron sea" of a metallic electrode and the "bendable" energy bands of the semiconductor. In this context, the instabilities are interpreted either as the unwanted carrying of holes and electrons across the two high, interfacial potential-energy barriers by tunnelling or carrier excitation; or the internal drift of various ionic species, often sodium or hydrogen. The transport medium can often be regarded as very pure fused silica, which allows us to use the large fund of knowledge of the physics of crystalline and fused quartz — makers of science policy might note that this is a good example of the benefits imparted to technology by apparently unrelated basic research.

The understanding of charge movement in silica may seem to be a somewhat esoteric field; but, for obvious reasons, it is increasingly important to convey this knowledge to electronics undergraduates and to the growing numbers of workers entering the broader field of the development of "electronic systems on a chip". This book fills such a need. It is, for example, the first time I have seen a large and carefully selected set of references on instability research.

Such research is carried out using a limited number of specialized electrical and optical techniques. These are outlined in 30 concise pages. The theoretical principles required include those of defect solid state physics, surface physics, lattice dynamics, conduction physics and molecular quantum mechanics, and treatment of these concepts is woven into the subsequent chapters. Among the research findings discussed are recent developments in surface states, mobile ions, polarization, hole trapping, electron trapping and electrical breakdown.

For many of the intended audience the book will be tough going. Readers are provided with reference to a few of the best, general solid-state technology textbooks, but the style of writing is somewhat flat and the graphics are unimaginative, coming, in the main, directly from research papers or other textbooks. Nonetheless, the author does address all of the important conduction and trapping effects which determine whether or not a given MOS device will be "good".

The exploitation of conduction and trapping effects in the MOS system is only just beginning, and the practical aspects of instability research reinforce the usefulness of a review such as that provided by Dr Davis. The use of new and more subtle effects in MOS devices not only ensures "good" (i.e. stable) devices, but also leads to ones which are "better" than those of rival makers. In this context, it must not be forgotten that microelectronic devices are used not only in parlour games but also in machine tools, racing cars and missiles. The country which possesses the fastest and best microcircuits will perhaps win not only commercial success but also more explicitly gladiatorial contests. □

Andrew Holmes-Siedle is Consultant to the Physics Department of the Fulmer Research Institute, Stoke Poges, Buckinghamshire.

● Another recent textbook on MOS devices is Oliver McCarthy's *MOS Device and Circuit Design.* The book, primarily aimed at students of electronic engineering, but also intended for practising engineers, is published by Wiley at £19.50, $44.95.

Statistical expertise on the shelf

P.T. Saunders

Biometry: The Principles and Practice of Statistics in Biological Research, 2nd Edn. By Robert R. Sokal and James F. Rohlf. Pp.859 plus tables. ISBN 0-7167-1254-7. (W.H. Freeman: 1981.) £19.50, $31.95. Tables hbk £16.30, $22; pbk £6.95, $9.95.

FOR some years now, Sokal and Rohlf's *Biometry* has enjoyed considerable success as a statistics text for biologists. Knowing this, my first impressions of the second edition were less favourable than I had expected. While the opening chapters are written in an attractive enough style and deal with the right sort of material, I found them too long and diffuse. The authors seem so intent on ensuring that every aspect of the subject is covered that they leave the reader in danger of not seeing the wood for the trees. To give just one example, the crucial idea of the testing of hypotheses seems to get lost in a long section devoted mainly to Type I and Type II errors and power functions. It is not that the latter are not worth discussing, though they are relatively less important in a text in which derivations are omitted, but that the balance is wrong.

But while I have some reservations about the introductory material, the remainder of the book, which consists of detailed accounts of a large number of statistical techniques applicable to a wide variety of problems, is very good indeed. The authors make excellent use of boxes, set aside from the main text, in which they show precisely how each test is to be carried out. They provide sample calculations for the benefit of those readers who are uncomfortable with subscripts and summation signs; more mathematically minded users will find them invaluable for testing computer programs. There are also informal discussions of the ideas behind the tests and other relevant points.

A number of improvements have been incorporated into the new edition and some additional material has been included, mainly concerned with correlation and regression. Inevitably the coverage is still not complete; for example, the authors have judged that their intended readership will be able to manage without such topics as sequential testing and Bayesian techniques. They have also chosen to publish the statistical tables separately instead of binding them with the book, which has its advantages so long as you can keep the two together.

Biometry would not be my first choice for someone with no previous knowledge of statistics, though it is certainly well worth considering. I would, however, strongly recommend it for anyone who already knows a bit about the subject and now finds himself in need of something beyond what he was taught. It is not quite as good as having a friendly statistician across the hall, but it can come surprisingly close. □

Peter Saunders is a Lecturer in Mathematics at Queen Elizabeth College, University of London.

nature

10 June 1982

Colleges back from the dead

United States universities and colleges are celebrating their commencement days with relief. The prospects for the next academic year seemed even worse a few months ago. But they are still grim.

In *Tom Sawyer* the hero attends his own funeral, only later to reveal to everyone that he wasn't dead after all This spring US colleges and universities are in a situation very much like Tom's For most of the year, the death of American higher education has been lamented by the higher education lobby in Washington, fighting the cuts in student loan finance proposed in President Reagan's budget Now, however, it looks as though Congress has heard the cries of middle-class parents throughout the country, and will rescue the education monies so that the young people who need government loans to pay tuition bills can go on attending college in the fall So higher education is alive, after all

But to be alive is not to be well A major debate has been going on in the US education community about the changing shape of US higher education including, literally, the ability of some kinds of schools to survive the next decade The month of June, when most schools' terms end, when final revenue figures are known and re-enrolment data are available, is a good time to take stock Moreover the blistering fight over government student loans, which has preoccupied university spokesmen, the Administration and the Congress for months, is now winding down The smoke is clearing And everyone agrees that US colleges and universities are under unprecedented stress

The most obvious source of stress is the declining number of 18-year-olds in the US population projected from the early 1980s onwards As a study by Princeton's president William Bowen said a few years ago, this could presage a general shrinkage in the size of the US university system Applications figures for enrolment in the fall of 1982 are down a mite (0 9 per cent), but educationists attribute this decline to other factors, such as rising tuition fees, high interest rates on private loans and the uncertainty of government loans The population factor will show up, they say, next year, when the number of applicants for the fall of 1983 drop off Eventually, the drop will mean that schools will have to compete harder to get students (or lower their standards) To anticipate this prospect Bowen recommended that the universities should plan to shrink

A second source of stress is the financial crisis in state governments, due to the taxpayers' revolt that spread to many states after it began some years back in California, and the loss of revenue from traditional federal programmes under the Reagan Administration Some states, such as Michigan, Missouri and Minnesota, simply cannot meet expenses any longer, are running enormous deficits and so are cutting back funds to their state universities State university financing has become an annual roller-coaster that would give ulcers to the average modestly comfortable professor in Cambridge, for example Some state systems have grown, over the years, into large, comprehensive undergraduate and graduate programmes, of high quality and easy access through nominal tuition fees They fulfill the best ideals of American education But some states have laws requiring that the budget be balanced every year So, for most of the year, university administrators do not know how much money will be given by the legislature for the next year, or whether the university will be the target of some last-minute cut

The easiest response is to conserve fuel, cut down on office staffs and maintenance, defer new building or expensive new computers or other instrumentation, to shave the number of faculty evenly across the board — and to avoid fights But that does not maintain quality To maintain quality a school must cut out the fat (weak departments) or do away with tenure — both of which can lead to embittered fights on campus and also lawsuits Another course is to raise tuition fees at state schools above the traditional nominal level (freshmen at the University of Southern California next autumn will have to pay $7,000, for example) But while this raises revenue, it discourages applicants and contradicts the very idea that state schools are meant to embody high quality education for everyone None of these remedies is pleasant, let alone conducive to an unperturbed environment for study

Not all state schools are in this same plight Some, such as Pennsylvania State University, are doing well But these are schools with strong ties with local industry and a well-understood tradition of service to the state and the community Applications for next autumn for Pennsylvania State University, for example, are up 10 per cent Industry finance is already a school tradition, so that door is open And so long as the cost of private and other state school tuition keeps rising, prospective students shopping for the best deal will come

The stresses on private schools are variations on those afflicting the state schools They cost more, so applications for next fall are down at most of them, according to a survey by John Minter Associates for the *Chronicle of Higher Education* The swing of the marketplace is clearest in the west and south-west, where 67 per cent of private schools report decreases in applications while 71 per cent of the public schools report increases Hardest hit of the private schools are the mid-ranked colleges, particularly liberal arts colleges, which cannot advertise quickly usable skills to the career-minded young and which must charge substantial tuition fees The choice facing these schools is either to lower admissions standards or "planned shrinkage"

Meanwhile, the rich get richer The top private schools are doing well, although they face several shifts in student choices and faculty availability The top schools often have large and important research laboratories supported by industry and the government, or major graduate schools or medical schools that are sources of overhead to help carry along departments, such as the liberal arts, that do not generally have outside income Part of the new, conservative public mood in the United States, as Chester E Finn Jr has written, is a concern with educational quality And, when parents, students, alumni and industry are all preoccupied with quality, they will tend to gravitate to institutions of proven quality Thus, Harvard's tuition fees have never been higher (next fall they will be $8,195) and its fund-raising has never been more successful

What does all this mean for science? It means that the trends in academic science identified in 1977 by Bruce Smith and William Karlesky* are probably going to be even more pronounced Smith and Karlesky surveyed many different kinds of US colleges and universities — four-year community colleges, state schools and top rank research institutions They identified about four principal trends, which would combine, they said, to squeeze out the "middle" of US academic research in the future At the lower level, the vocational schools would prosper (as, indeed, they have) because when people are unemployed, they go to school to get

*Smith, B & Karlesky, W *The State of Academic Science* (Change Magazine Press, Washington D C , 1977)

0028-0836/82/230443-02$01 00

skills But as funds become tight, mid-ranked schools which can compete less well for federal funds, industry gifts and students will become less and less able to carry out a research as well as a teaching role Their research programmes will diminish Meanwhile, the top-ranked schools, while facing critical problems such as a decline in places for young faculty, will ride out the storm Like Tom Sawyer, they will live on

New ways with weapons

Bilateral negotiations on strategic arms are to begin later in the month. Where will they lead?

The passage of time has curious healing qualities Two years ago, Mr Ronald Reagan, campaigning successfully for the presidency of the United States, was outspoken in his condemnation of President Jimmy Carter's second instalment of the treaty that emerged from the Strategic Arms Limitations Talks (Salt II), originally devised by Dr Henry Kissinger for once-President Richard M Nixon Reagan was pushing at an open door, for his political opponent knew that the US Senate would not ratify the treaty he had signed, indeed, the Soviet occupation of Afghanistan gave him a handy excuse to withdraw his request for ratification without too much loss of face The Senate's objections to Salt II were largely technical — verification, the equity of the equation of one long-range missile with another and, perhaps more important, the knowledge that nobody from the Carter White House had faced the problem of persuading the Pentagon that the rules of the strategic arms race should be changed Candidate Reagan could afford to be more radical and rhetorical how is it possible to trust in an agreement with people whom we do not trust? President Reagan has nevertheless found that his administration has kept within the unratified rules of Salt II, if only because it has been hard to know what should be done about the disposition of the MX missiles, the long-range missiles destined to replace the Minuteman In the Soviet Union, likewise, the same rules have adventitiously been kept And now, the emollient passage of time has persuaded President Reagan that he must follow his predecessor's path and propose bilateral talks with the Soviet Union on strategic arms (see *Nature* 20 May, p 169) The surprise is that the Soviet Union has so quickly agreed the talks are due to begin in Geneva on 27 June

A modicum of muck-raking is in order even in these hopeful circumstances The Reagan Administration's original indifference to strategic arms talks appears to have been largely ideological — sup with the devil only with a long spoon, negotiate on strategic arms only from strength This position has been undermined by predictable (and predicted) forces — political pressure from allied governments, political pressure from domestic voters, fears that another review conference of the Non-Proliferation Treaty (in 1985) as arid as that of 1980 might drive some middle-sized powers into the nuclear club, the sheer cost but also the technical difficulty of developing the new generation of strategic arms and the recognition that the world, with the Falklands and the Middle East, is a dangerous place

So what should the Reagan Administration be aiming for? Its opening bid is simple enough — let each side be constrained within a limit of 5,000 strategic warheads with the proviso that no more than half of these should be based on land The proposal is rational enough — submarine missiles are for the time being relatively invulnerable to attack, while ground-based missiles are potentially vulnerable to a first strike by the other side and are thus destabilizing The Soviet Union's predictable but far from final objection is that the proposal would more immediately affect the Soviet Union than the United States The success of the Geneva enterprise will crucially depend on the willingness of the two sides to acknowledge that 5,000 nuclear warheads, each of them at least ten times larger than the bombs dropped on Hiroshima and Nagasaki, are outrageously too many for the now classical purposes of deterrence Indeed, these huge numbers can be explained only if each superpower is prepared to attack the other's ground-based missiles in a supposedly pre-emptive first strike, in the process presumably inviting an attack on its cities by the other's submarine missiles Are the two superpowers prepared to return to the simple view that nuclear weapons are intended simply as deterrents?

Other difficulties will arise before that question is faced How, for example, will the new strategic arms talks be linked with the talks on "theatre nuclear weapons" already under way in Geneva? That the two negotiations should be linked is sensible, while the United States proposal that its FB111 bombers based in Europe should be counted within the separate limit for strategic aircraft is an important concession But where will the British and even French nuclear forces fit in? And would the attainment of President Reagan's "zero option" for European "theatre weapons" be the equivalent of the nuclear-free zone for which the Palme commission was pressing last week (see page 446)?

These questions are so difficult that there can be no quick end to the proceedings in Geneva The best hope is that there will be some quick agreement on a warhead ceiling below which further reductions will be sought During that second stage of the negotiations, the Soviet Union will be faced with the theoretically sensible but politically sensitive proposal by the United States that limits on strategic weapons should be determined not by counting missiles or even warheads but by their throw-weight of carrying capacity The obvious snag is that this criterion would more severely penalize the Soviet Union (with its large rockets) than the United States, while everybody would know that the effective power of a warhead is at least partly a function of its accuracy, quantifiable but hardly verifiable This is why the most important potential benefit of the Geneva process is that it may change the way the superpowers think — and that it might in particular persuade people like Mr Alexander Haig, the US Secretary of State, to avoid statements like that the other day in which he said that no strategic arms agreement would be acceptable if it made conventional wars more likely Is that a sensible criterion? Nobody thinks conventional wars desirable in themselves, but an effective agreement on strategic nuclear weapons must surely be one that will allow conventional warfare to break out without the immediate risk of nuclear conflict

Waiting for Merrison

British university research is threatened by attrition and neglect. The Merrison report will not help.

After more than two years of brooding, Sir Alec Merrison's committee on the condition of British university research seems to have confirmed people's worst fears but to have failed to show how they may be exorcised Its report just published (see facing page) is a masterpiece of judiciousness The dual support system whose collapse in the past decade was the chief reason for the committee's existence is said to be in all ways admirable, and therefore is to be preserved at all costs Indeed, to judge from what the committee says, there is only one defect in this excellent and peculiarly British device for sharing the cost of university research between one public purse (the University Grants Committee) and another (the research councils) the system appears not to function as intended In much the same spirit might a man sent to rescue a motor-car fallen in a river report that the car was in excellent condition but that it had sunk

Nobody, or hardly anybody, is to blame for this debacle The committee had the bad luck to have been established before the full severity of the British government's intentions towards its universities had become apparent And the committee's terms of reference (behind which it shelters) were in truth too narrow It was allowable to observe (as the committee does) that universities' internal arrangements for deciding priorities in their discharge of responsibility for research and scholarship are inadequate, but outside its brief to enquire why things have come to such a pass The more serious difficulty, however, is that British universities are in such a mess, one so arbitrarily brought about, that the safeguarding of research also requires arbitrariness on somebody's part Only the research councils are at present capable of being as selective as circumstances require

0028-0836/82/230444 01$01.00

Nuclear waste bill now in sight

But critics fear stop-gap stores will last

Washington

Congress is now closer than ever to passing a nuclear waste management bill, but its terms seem likely to be much more appealing to the atomic industry than to environmental groups, both of which, in a strange alliance, have for years been calling for a legislative solution to the growing mass of spent commercial fuel.

The Senate has already passed a bill that the industry is satisfied will end the uncertainty plaguing the government's waste disposal programme. In the absence of legislative directions, each administration has been free to set its own policy, usually inconsistent with that of the previous administration.

The environmental lobby, on the other hand, is worried that the Senate bill and its counterparts now under consideration in the House gloss over the serious technical problems of waste management in favour of political expediency.

At present, 8,000 tonnes of spent commercial fuel is in temporary storage at reactor sites. By the end of the century, the figure is expected to reach 72,000 tonnes. The ultimate solution, everyone seems to agree, is to dispose of it in deep geological repositories. This solution is provided for in all versions of the bill.

The bone of contention, however, has become whether the federal government should in addition provide some form of interim storage. According to the Atomic Industrial Forum, which represents the nuclear power industry, roughly half a dozen reactors will run out of on-site storage space by 1985, the problem will be widespread by the 1990s. The industry is thus very pleased that the Senate bill provides for stop-gap storage to cover any delays in a permanent repository.

The environmental groups see something more sinister going on. The interim storage envisaged in the Senate bill will be of two kinds. The first is "away-from-reactor" (AFR) storage, which is essentially the arrangement used at reactor sites: spent fuel elements are simply stacked in a water-filled "swimming pool" which absorbs the radiation and heat. The environmental groups charge that AFR is a way for the industry to avoid the licensing procedure required to expand on-site storage. The second kind of facility is "monitored retrievable storage" (MRS), which is only vaguely defined. The chief worry among the environmental groups is that MRS will become the *de facto* permanent solution.

Brooks Yeager of the Sierra Club says that the Senate bill virtually guarantees that. It sets "not just ambitious, but unmeetable deadlines" for the construction of a geological repository. "They want to arrive at the issuance of a construction permit by the end of the decade. That's seven years faster than the Department of Energy's plans for construction in order to resolve all the technical problems." The timetable may also guarantee that the choice of sites will be limited to the three at which the Department of Energy has already begun tests — the Hanford Reservation in Washington state, the Nevada test site and a group of sites along the Gulf of Mexico.

Professor Henry Kendall of the Massachusetts Institute of Technology, who is active in the Union of Concerned Scientists, a group critical of US nuclear policy, agrees that building an MRS facility "basically means you don't have confidence in permanent disposal". He says that what is needed is time to make a careful hydrological study of the actual site.

These worries seem to be backed up by a recent study by the Congress's Office of Technology Assessment (OTA). The

British research — no cure yet

British university research is in trouble, but only the universities themselves can work their way out of it. This is the chief conclusion of the much delayed report of the committee under Sir Alec Merrison, vice-chancellor of the University of Bristol, set up two and a half years ago to brood about financial support for university research and published this week (Cmnd 8567, HMSO, £4.35).

The report comes down squarely in favour of the British dual-support system, whereby the University Grants Committee (UGC) provides universities with the basic wherewithal for research and the research councils provide extra funds (but no overhead) for particular projects. But the committee also says that "the system has been under strains for several years".

One of the committee's chief proposals is that universities should more deliberately channel part of the funds they receive from the UGC into areas of research in which they consider themselves to be strong, for which purpose it says that British universities should set up research committees to supervise the internal allocation of funds. But the committee also says that as a stop-gap, the research councils should be prepared to "meet costs they would not normally meet" (a euphemism for paying overhead) or think of moving people doing good work in unfavourable circumstances to other places.

The essence of the committee's support for the continuation of the dual-support system is its repeated reaffirmation of the belief that if universities were not provided with funds that can be spent on their own discretion on research projects that would not normally win research council support, genuinely innovative ideas would never see the light of day. It considers but rejects on the same grounds that UGC support for research should be linked directly with the volume of financial support provided by the research councils, while it considers that if UGC were to earmark any but a small proportion of its university support for specific projects, the resulting rigidity would be self-defeating.

The chief targets for the committee's advice are the universities, which are told that in the long run — "the prospects for achieving any significant shift in the near future are next to impossible" — they must be prepared to spend more of their resources on research rather than teaching, that they should "concentrate research funds into selected areas", look at the problems occasioned by academic tenure, find ways (with the help of the research councils) of bringing in "new blood" and be prepared to form associations with other universities for more effective prosecution of research. Both partners in the dual-support system are asked to be more sensitive to researchers' need to travel.

The research councils are given two principal tasks — to adjust the support provided for graduate students more regularly in tune with the increasing cost of living and to "study" the balance between their support of research in universities and in their own establishments. In a memorable sentence, the report says that "we are not satisfied that the balance of research council expenditure between such support and the work of their own institutes is in all cases right".

The committee's belief that university research is in trouble is based on statistical evidence that the committee says should be improved. The data do, however, show that the decline in research support from universities' own budgets goes back to the early 1970s, and that between the beginning and the end of that decade the average sum of money available to university departments per head of academic staff employed declined by 28 per cent in real terms.

These and supporting figures appear to have prompted the only note of near-acerbity in the committee's report — that while the British Prime Minister has repeatedly stated that the science budget has been protected, "the health of university scientific research does not depend only on the science vote". ●

0028-0836/82/230445-02$01.00

report* urges that waste be kept at reactor sites until a permanent repository is built to "avoid diverting the attention and efforts of the waste management agency away from the repository program toward provision of an independent interim storage facility"

Another sore point for environmental groups in the Senate bill is an amendment that Senator James McClure succeeded in attaching, which declares that the legislation itself represents reasonable assurance that a safe disposal method exists Yeager says "It's a clear attempt to end-run several court cases and state laws" California, Oregon and several other states have passed laws restricting new power plants until such reasonable assurance exists

To the industry, however, the amendment is nothing more than recognition of what they see as the obvious the waste disposal problem is political, not technical And here, the Office of Technology Assessment report backs the industry view It found "no insurmountable technical obstacles" to development of geological repositories, rather, the chief obstacle is eroded public confidence, aggravated by a vacillating federal policy President Carter, for example, reversed the previous policy of handling defence and commercial wastes separately, President Reagan reversed the Carter policy Presidents Reagan, Carter and Ford each changed the number of sites under study for geological repositories and their construction schedules

If a bill does emerge from Congress this year, it will almost certainly contain provisions to end this instability Both Senate and House versions provide for long-term funding through a surcharge on nuclear electricity, which could raise $300 million in the first year

Both Senate and House versions also set up mechanisms for state participation in the site-selection process The federal government's insensitivity to the concern of the states in previous siting decisions is viewed as a major factor in the loss of public confidence in the programme

The chances that Congress will act this year are, according to all parties involved, better than they have been in past years In 1980, both the Senate and the House passed bills, but failed to work out a compromise This time, they start closer

According to a staff member of Morris Udall on the House Interior Committee (Udall introduced the House bill), "there's no reason we can't get a bill except time" The prediction is that if the bill reaches the full House by mid-July, it should pass

But passage may ultimately hinge on the complex politics of the House committees No fewer than three have asserted jurisdiction over the measure, which is now tied up in the Energy and Commerce Committee **Stephen Budiansky**

Managing Commercial High-Level Radioactive Waste (Office of Technology Assessment, May 1982)

International disarmament

Palme's palm

A nuclear-weapons-free zone in Europe and greater United Nations (UN) power to prevent hostilities in developing countries are the chief recommendations of the Independent Commission on Disarmament and Security Issues whose report, *Common Security a programme for survival,* was published last week Although neither recommendation is new, Dr David Owen MP, former British Foreign Secretary, and Sir Shridath Ramphal, Secretary-General of the Commonwealth Secretariat, who presented the report in London, argued that their revival now is timely

So, partly by design and partly by accident, was the publication of the report The original intention was to produce recommendations for the UN Special Session on Disarmament which began this week But the commission now also hopes that its document will inform the Strategic Arms Reduction Talks (START) due to begin in Geneva on 27 June Dr Owen is optimistic that the South Atlantic conflict will focus attention on the recommendation for a UN fact-finding mission which could be called in by a country fearing imminent aggression by a neighbour

The report comes after more than a year's deliberation by 17 distinguished politicians from as many countries under the chairmanship of Olof Palme, former prime minister of Sweden The task of the commissioners, each of whom was chosen for their understanding of international affairs and not as a national representative, was to prepare a report on the consequences of the arms race along the lines of the Brandt report on economics

The result is an analysis of global conflict and tension which embraces the effects of nuclear arms build up on North–South as well as East–West relations and the effects of the increasing sophistication and strength of conventional forces on relations between developing countries The commission believes that the risk of a devastating war is increasing in spite of attempts at arms limitation It says that national security can no longer be assured by military means

The idea of a limited nuclear war and the conventional strategies of the NATO and Warsaw alliances on the early use of strategic nuclear weapons are challenged The commission's answer is to create a nuclear-weapons-free zone in Europe by negotiating parity in conventional forces at reduced levels

Over the past eight years, the report says, developing nations have been drawn into the arms race, while vertical proliferation in the five recognized nuclear powers has increased the probability of horizontal proliferation Industrial nations have also made matters worse by selling sophisticated conventional arms to developing countries, making the escalation of local disputes more likely The commission is particularly critical of the high level of military spending by nations that can least afford it Security for developing nations, particularly those with small populations whose borders are in dispute, must be assured by other means — hence the recommendation to revitalize the original intention of the UN charter by increasing the powers of the Security Council to pre-empt conflicts

The commission's "programme of action" is divided into short and medium-term measures, for implementation within two and five years respectively The programme includes measures to reduce nuclear and conventional forces in the NATO and Warsaw alliances, a nuclear-weapons-free zone, a comprehensive chemical weapons disarmament treaty, agreement on guidelines for conventional arms transfer, a substantial reduction in military spending and the transfer of military scientists to civilian research

Little advice is offered on how to achieve these measures But the commission is clearly hoping that the chief impact of its report will be to focus increasing public concern **Judy Redfearn**

Italian science policy

New consensus

Venice

Italian science policy appears at last to be getting into gear, despite the almost annual changes of government that confound most long-term planning in this country The reason? Politicians here have reached a consensus close to the French position that research is necessary to pull Italy out of its economic crisis

This consensus is beginning to survive the rise and fall of governments, so the present science minister, lawyer Giancarlo Tesini, who has been in post for a year now in President Giovanni Spadolini's shaky five-party coalition, has been able to take some action

One of his key moves has been to achieve five-party agreement on a reform of the Italian national research council (CNR) which until recently has dominated Italian government research and development both inside and outside universities "CNR is a major problem of the Italian science ministry", Tesini said here last week, where he is attending a scientific meeting

Broadly, Tesini wants to shift CNR, with most of its laboratories in universities, into a closer relationship with industry The reform would put industrial scientists into key CNR positions and streamline an organization which is widely regarded as being massively bureaucratic — so bureaucratic, in fact, that in 1980 it failed to spend an important part of its budget concerning the Italian space programme For its pains, the result in 1981 was a budget cut of 10 per cent in current lira (much more in real terms) to the present figure of

0028-0836/82/230446-02$01 00

around 420,000 million lira (£180 million)

Under the previous government, CNR also lost control of university research (that is, research in universities outside CNR laboratories), which is now in the hands of the national council of universities (CUN), where the same professional groups take similar decisions but outside the CNR bureaucratic structure

Tesini's reforms, now before parliament, would revitalize the crumbling structure of CNR Already CNR is spending an increasing part of its budget on applied projects, the latest list of seven amounting to 300,000 million lira over five years (Two examples of these projects are an effort to increase Italian agricultural productivity, particularly in the south, since Italy is a nett food importer, and a recently completed project to develop an electronically-controlled fuel-efficient engine in collaboration with Alfa Romeo)

But Tesini does not want to make CNR entirely into an organ of industrial research "It would be dangerous to create a gulf between university work and CNR" he says

Now the future of CNR is in the hands of parliament, but in the nature of Italian politics it may be two weeks or a year before the vote is taken on the Tesini reforms, by which time his government may well be out of office and a new minister in place to take the credit **Robert Walgate**

Private help

Washington

In a move that will help fill the void left by the Reagan Administration's cutbacks in environmental research, a private group — the MacArthur Foundation — has stepped in to offer $15 million for a new research organization The Institute for World Environment and Resources will concentrate on issues in global ecology such as species loss, inadvertent climate modification, desertification, deforestation and population

A key government agency that had been following long-term global issues, the Council on Environment Quality (CEQ) was cut back to a bare-bones staff as soon as President Reagan took office The director of the new institute, James G ("Gus") Speth, was chairman of CEQ under President Carter

The institute, which expects to begin operations by October, will however attempt a broader and more scientific approach than CEQ has taken It seeks to bring together scientists from the physical, biological and social sciences to work on interdisciplinary problems of significance to public policy

Although the institute will be based in Washington, roughly one-quarter of its $4 million annual budget is expected to go to established research contracts with university centres of expertise in the environment and resource fields

Stephen Budiansky

Humid tropics

Cautious growth

Washington

In a report released last Thursday, the National Academy of Sciences has recommended cautious but expanded development in the humid tropics The study, commissioned by the US Agency for International Development*, found that concern over the environmental consequences of tropical deforestation is legitimate, but often overstated, and moreover that development is inevitable Ninety per cent of the world's population increase from now to the end of the century will take place in the tropical countries — an addition of 1,500 million people

One of the main messages of the report is that despite the inherent difficulties in managing tropical areas — and the study recites them all, including annual rainfall of 5–35 feet, tremendous species diversity, and poorly understood ecology — new technologies are appearing that can get around these problems

According to Dr Pedro Sanchez of North Carolina State University, a member of the panel that prepared the report, some myths about tropical soils need to be dispelled He points out that "there's nothing fundamentally different in the agronomy between the humid tropics and the southeastern United States Contrary to myth", he adds "tropical soils though often infertile and acid are not fragile They are not especially low in organic matter nor especially prone to erosion A lot of the erosion pictures in the popular press are 'civil engineering erosion' pictures taken around roads and drainage systems"

Sustained agriculture on cleared land does, however, require careful management and new technology, in particular continuing application of fertilizer Sustainable agriculture, as opposed to traditional shifting cultivation, is considered the key to attaining the greatest benefit with the least deforestation and thus the least environmental disruption

Still, much may be possible through the fine-tuning of indigenous technologies For example, the traditional slash-and-burn method of clearing turns out to be much easier on topsoil than are bulldozers, and furthermore provides free fertilizer in the form of nutrient-rich ash

Dr H Garrison Wilkes of the University of Massachusetts, Boston, another panel member, points out that too little has been done to exploit and improve on tropical germ plasm resources "The amount we know about cowpeas," he says, "or other tropical crops such as casaba, palms, tropical cabbages is very small — so the potential for genetic improvement is tremendous " **Stephen Budiansky**

**Ecological Aspects of Development in the Humid Tropics* (National Academy Press, Washington DC, 1982)

Soviet agriculture

Food for all?

The Soviet Union's new food production programme designed to make Soviet food production match demand was announced last month at a plenary meeting of the Central Committee and introduced in a major speech by first Secretary Leonid Brezhnev himself

To avoid emergency purchases from the non-Socialist world while this drift from agricultural to industrial employment continues will demand a major investment in agricultural research Several lines of investigation are specifically mentioned in the programme, but their implementation is formally entrusted to the leading Soviet scientific bodies — the Academy of Sciences, the Lenin Academy of Agricultural Sciences and the State Committee for Science and Technology

According to the programme, however, the practical work is to be the responsibility, primarily, of the network of "scientific production associations" set up in the past few years to link research and development organizations with production enterprises These associations are to form the basis for the production of high quality and hybrid seeds and plants of "superior reproductive quality" as well as the rearing of pedigree cattle for state and collective farms

The targets laid down by the new programme are precise varieties of winter wheat must be developed with a yield of not less than 80–90 q ha^{-1}, spring wheat of 45–60 q ha^{-1}, maize hybrids of 12–130 q ha^{-1} on irrigated land and 80–90 q ha^{-1} on non-irrigated land and peas of 40–45 q ha^{-1} (1q = 100 kg) Fodder crops must yield 10,000–15,000 fodder units per hectare under irrigation and 5,000–6,000 fodder units without, while technology or fodder storage must ensure retention of at least 90 per cent of nutrients

In addition to these precise targets, there are less specific directives Research is to be directed at saving energy in soil conservation, the rational usage of water resources, anti-pollution measures and the mass examination and treatment of animals Special attention is to be paid to the development and mass production of biological and chemotherapeutic veterinary pharmaceuticals

To support this applied research, the two all-union academies, together with the republic academies of sciences, are to develop theoretical work in genetic engineering This will include the breeding of new strains of plants, microorganisms and animals, the biotechnology of protein synthesis and biologically active substances The academies will also be responsible for new pesticides and herbicides, growth regulators and similar sophisticated agricultural preparations, and for designing the industrial technology for their commercial preparation **Vera Rich**

0028-0836/82/230447 01$01 00

US National Institutes of Health

New clinical trials

Washington

Dr James B Wyngaarden has finally been sworn in as director of the National Institutes of Health (NIH), the US health research agency, and so has come out of the shadows where he has been hiding, in a sense, since his name surfaced as the Administration's candidate for the post last autumn He has now begun interviewing candidates for those institutes that have no permanent directors at the moment (some of the posts have been vacant for months)

Last month Wyngaarden also held his first press conference in his new capacity, and the debut produced one serious slip — a statement that his personal belief favoured freedom of individual choice regarding abortion — which was immediately attacked by anti-abortionists, who are very active in Washington these days He also said, speaking personally, that he would have "no problem" with carrying forward NIH's research in animal *in vitro* fertilization to the human level Whether NIH should sponsor research on human *in vitro* fertilization has been a controversial subject — so controversial that possible government guidelines have been pending, but not issued, since the time of the Carter Administration

But when he spoke officially, Wyngaarden indicated his views for the future of NIH — which he has had ample time to work out in the long months of warm-up for the job He said he wants to keep the present balance of activities, particularly the level of 50 per cent of NIH funding allocated for basic science including investigator-initiated grants He also favours continuing use of training grants and fellowships as a means of producing good basic biomedical scientists He said he agreed "fundamentally" with the stabilization programme meant to insulate NIH from budget swings that was introduced by his predecessor, Donald K Fredrickson Wyngaarden also defended peer review as a "brilliant creation" of one of his predecessors, James A Shannon, and said that NIH were not at the moment well placed to embrace more institutes All this was hardly surprising, given Wyngaarden's background in academic medicine at Duke University School of Medicine

Wyngaarden did, however, express interest in starting new clinical trials NIH still run several, but some have come under criticism for being very expensive and sometimes not useful No new clinical trials have been initiated by NIH since 1978 But Wyngaarden said he was considering suggestions for new clinical trials in the cancer, cardiovascular and arthritis fields that might well be worth starting even though they would commit funds for many years and take money away from investigator-initiated grants New clinical trials would, of course, mean a change of the balance of activities of NIH But such a change, if it comes, is some time off, as Wyngaarden has much to do now to get NIH running again **Deborah Shapley**

Biotechnology

Yugoslav plan

Zagreb

Yugoslavia is planning a major drive in biotechnology, based on a projected new institute on the Dalmatian coast The new scheme, which now awaits federal approval, has the backing of the government of the Croatian Republic, and is intended both to enhance the international prestige of Yugoslav science and to help the drive to develop science-based industry

The present stagnation of the Yugoslav economy has made it impossible for science and industry to continue its recent course of dependence on foreign licences and the import of obsolescent technologies Indeed last year's Congress of Self-Management stressed unequivocally the need for home-grown technology

Some research institutes embarked on such a course several years ago Thus the Jozef Stefan physics institute in Ljubljana supplies knowhow to a range of industries from medical technology to nuclear power and the Immunology Institute of the University of Zagreb has become a major earner of hard currency by means of sales of sera and sophisticated pharmaceuticals, including interferon

The new emphasis on science-based industry, however, accords well with the desire of Yugoslav — and particularly Croatian — biologists to see their country take a lead position in biotechnology During the academic year 1979–80 a group headed by Dr Marija Alaceivic of the Department of Biotechnology of the University of Zagreb raised the possibility of founding an international institute for molecular genetics somewhere on the Croatian coast

Several coastal towns were asked to tender for providing a site Dubrovnik, which already boasts an Inter-University Centre of Postgraduate Studies, was tipped as favourite Just before the closing date, however, Split made an intriguing counter-proposal The institute, Split suggested, could be housed in the Villa Dalmatia, one of the residences of the late President Tito The proposal now only awaits official ratification Apart from symbolic implications, Split has other advantages including a long-standing academic tradition which, in 1974, culminated in the foundation of a university

The choice of biotechnology seems to be dictated partly by economic considerations — the costs involved in biotechnology, while not negligible, are modest Moreover, Yugoslavia has a long tradition in organic chemistry and can boast two Nobel prize winners — Ruzicka and Prelog — in this field

Supporters of the institute already speak of the projected institute as "international" There is hope that the European Molecular Biology Organization (EMBO) will help but that would mean that Yugoslavia would first have to join EMBO — a proposal already formally made by the Yugoslav scientific community to the federal government Yugoslav participation in international scientific research has not always been auspicious — although a founding member of CERN, for example, Yugoslavia had to drop out because it could not meet the financial contributions

According to Professor Ivo Slaus, the chairman of the steering committee for the institute, even if no international support is forthcoming, the Croatian scientific community would still wish the institute to be "international" in the wider sense of the word The Republic of Croatia has already allotted 15 million dinars (£188,000) for preliminary work and part of this sum will be spent on a conference next year to launch the new institute and to ensure that it does not duplicate the work of existing institutes (such as Szeged in Hungary) **Vera Rich**

UK–Argentine cooperation

Shadows of war

The storm clouds over the South Atlantic are casting long shadows over scientific relations between Argentina and the United Kingdom Scientific collaboration between the two countries had been growing but it now looks as though several promising projects may be blighted The rumblings have already caused the postponement of the Sixteenth International Symposium on Remote Sensing of Environment, which was to have been held in Buenos Aires this week Organized jointly by the Environmental Research Institute of Michigan and the Comisión Nacional de Investigaciones Espaciales, the conference was to have heard 200 papers When the Falklands crisis erupted, many intended participants withdrew their papers and gave notice that they would not attend the symposium

Scientific collaboration between Britain and Argentina — in the form of joint university projects, exchange visitors and fellowships — has been nurtured over the past few years by the Royal Society and the British Council The Royal Society and Conicet (its Argentine equivalent) have had a reciprocal agreement for some ten years — 110 scientists a year have been exchanged All this has now been put on ice The Royal Society recently sent out a letter to scientists it had planned to sponsor on visits to Argentina saying that, in view of present circumstances, all exchanges should be halted

The British Council has increasingly encouraged British academics to put out

0028-0836/82/230448-02$01 00

feelers and develop reciprocal research projects with Argentina. Through the council's office in Buenos Aires, Argentine universities would suggest projects for possible cooperation and the council would then approach the universities in Britain. Travel expenses were paid by the visiting country's organization, the council or Conicet, and subsistence by the host. The council has now closed its office in Buenos Aires and recalled its staff.

When the Falklands crisis broke, eighteen jointly-funded university projects were under way. In the past year the British Council has sponsored 7 long-term scholarships, 36 visitors and 17 lecture trips. The numbers may not be great but the Royal Society, the British Council and many academics have expressed their profound regret that the present situation may put a big question mark over such agreements.

Several promising projects have been nipped in the bud. Imperial College and the Centro de Investigaciones Opitcas (CIOP), La Plata, have for the past eighteen months been discussing the setting up of a joint laser physics research project. Professor G H C New of Imperial College recently paid a visit to Argentina to give a series of lectures and arrange future collaboration. As well as exchanging postdoctoral fellows on extended visits — beneficial for each party in providing highly specialist researchers free — it was also hoped that a British specialist would go out to Argentina to teach and act as a consultant at CIOP.

Professor J Barber (Pure and Applied Biology, Imperial College) has also recently been to Argentina to coordinate a joint photosynthesis research project. The oceanography department at the University of Southampton has been making moves to coordinate research with Argentina, particularly in the area of oil exploration. Professor A P M Lockwood has been closely involved in talks to set up training exercises particularly in his capacity as a delegate of TEMA (Training and Education for Mutual Advantage), part of UNESCO's International Oceanographic Commission.

Both Argentine and British scientists are concerned that the Falklands crisis will seriously damage opportunities for scientific cooperation. In particular, many Argentine scientists in Britain are worried that if they leave the country to attend conferences abroad, they may have problems getting back into Britain. In April, both Britain and Argentina introduced a visa requirement. A spokesman for the Home Office did say, however, that he could not foresee any problem if Argentine nationals applied for re-entry visas before leaving the country. What is clear is that many academics feel uncomfortable in a situation where politics is impinging on their research. As one who has recently visited Argentina said, "They're a wonderful country and wonderful people. It's very, very sad."

Jane Wynn

Australian telesatellites

Outback opened

Canberra

Australia will have its own domestic communications satellite by the end of 1985. In a statement to the House of Representatives on 6 May, the then Minister for Communications, Mr Ian Sinclair, announced that Hughes Communications International will build the space segment of a domestic satellite. The system, known as Aussat, will be owned and operated from the start as a commercial venture by Aussat Pty Ltd. Although Aussat is at present wholly owned by the federal government, the government intends to sell 49 per cent of its equity to the private sector as soon as practicable, in line with its free-market policy principles.

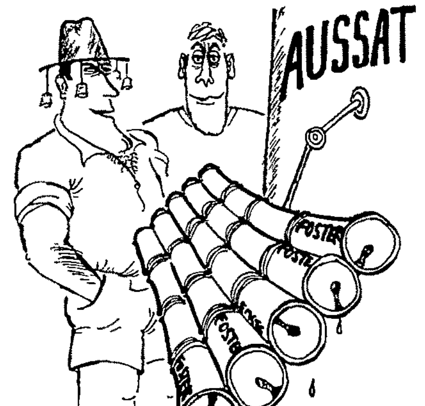
The contract with Hughes, the sealing of which seems imminent, specifies three satellites for the system — two in orbit and one on ground standby, with the option of a fourth satellite should demand justify it. The satellites will have an operational lifetime of seven years and function in the super-high frequency (gigahertz) range with four high-power and eleven low-power transponders. The net result will be a versatile satellite of limited capacity, primarily suitable for television networks. Provisional bookings with the European Ariane launch vehicle and with NASA's Space Shuttle or Delta rocket have been secured. The fixed-price package offered by Hughes, including the launching and construction of two ground control stations, will cost A$166 million (£95 million). Of this, only A$5.1 million will be spent on local technology, which is a sad reflection on the dearth of innovation in the Australian electronics industry.

Aussat promises a glittering array of services including direct television broadcasting to widely scattered homes throughout the continent, business communications and remote educational and medical services. Both the School of the Air, plagued by poor reception, and the Royal Flying Doctor Service will get a new lease of life. The satellite will also bring telephone services to the outback as well as a back-up service for trunk lines between major cities. Unfortunately, the more imaginative aspects of satellite telecommunications such as "teleconferencing", direct radio contacts and electronic news gathering are ruled out in this satellite because of its frequency allocation and limited number of transponders.

The principal government user of Aussat will be the Australian Broadcasting Commission, which has reserved all the four high-power transponders on the first satellite for its Homestead and Community Broadcast Satellite Service. The beneficiaries will be farmers and station-owners who can afford A$1,000 for the purchase of a 1.2-metre diameter reception dish. For the rest of Australian viewers, the satellite will bring more of the same, the Australian Broadcasting Commission is likely to increase the proportion of its programmes relayed in real time, rather than on film or tape, to its network. On the other hand, Aussat may facilitate the production of special-interest programmes for small groups such as ethnic communities, for which there is insufficient local appeal but which may be viable on a national basis.

Other public users will be Telecom, who will provide the remote telephony and business services, the Department of Transport, which will use the facility for aeronautical and marine communications and the state police departments.

The commercial television stations are expected to be the principal private users. There is, however, the possibility of using



"At least we made the solar panels"

the high-power transponders on the second satellite on a pre-emptive basis, in which case the fourth satellite will certainly be needed. Significantly, Aussat will provide the first opportunity for a private operator to compete with Telecom. At present, legislation ensures that Telecom has a monopoly in all landlines and terrestrial microwave links. However, Telecom has a statutory obligation to weight its charges for profitable services in order to cross-subsidize telephone services in rural areas. A competitor, not hampered by these obligations nor by having to contend with the growing union activism within Telecom, could under-cut Telecom for its business services via satellite and still make massive profits.

The substantial private equity in Aussat raises political questions of regulating ownership, access and frequency allocation. The satellite, by enhancing television networking, will undermine the position of commercial stations not having access to it. Consequently there is a real danger that the media, now already in a few private hands, will be even further concentrated. Another problem to be tackled is the possible piracy of the downlink beam by neighbouring countries or illegitimate access to international and overseas databanks. Australian legislators will have to resolve these problems with little by way of precedent.

Vimala Sarma

0028-0836/82/230449 01$01 00

CORRESPONDENCE

A history of the Falkland Islands

SIR — In your leading article referring to the Falkland Islands conflict (see *Nature* 8 April, p 480), it is stated that the only claim of Argentina to the ownership of the islands is the proximity to the Argentine coast Since we consider this statement to be in error we give here a detailed account of the historical events and the cause of the present conflict

The discovery: The strongest evidence available indicates that the discovery of the islands can be attributed to the crew of the ship "San Anton", a member of the Spanish expedition led by Hernando de Magallanes (Magellan) With the name of "Islas de Sanson" (an abbreviation and slight modification of the ship's name) the islands are depicted in several Spanish maps from 1522 to 1590, as well as in the Italian map of Agnese, of 1536, where the Magellan route is indicated

Naturally, Spanish dominion over the islands was proclaimed (in accordance with Papal bulls and treaties with Portugal) and in 1580 a garrison was settled near the Magellan Strait having jurisdiction over the mainland and the nearby islands

In 1619, on a map made under the direction of the Dutch sailor Sebald de Weert, the islands appear with the name of "Sebaldin islands" The first British descriptions were published in 1622 but even British scholars are doubtful about these reports

However, the British sea rover William Dampier was there by 1648 and in 1690 John Strong gave the name "Falkland Sound" to the strait between the two main islands, a name that in the English literature was later extended to the whole archipelago

The occupation: The islands were first occupied in 1764 by the Frenchman Louis Antoine de Bougainville who established a small colony in a place he named Port Louis Since most settlers were from the French port of St Malo the islands took the name of "Iles des Malouines", which in the Spanish literature became "Islas Malvinas"

Once Spain realized that the French had settled there the Crown made the appropriate claims to France and in 1767 the French left A small Spanish garrison subsequently remained in Port Louis, which took the name of "Puerto Soledad"

In the meantime (1765) a small group of people from Britain had settled in a place which they called Port Egmont Once again, Spain made the respective claims and since the British did not want to leave peacefully they were expelled by force in 1770 The diplomatic conflict that started ended by an agreement between Britain and Spain in which the British were reinstalled in Port Egmont in 1771 and left "spontaneously" in 1774 The British occupation lasted four years in all

In 1811, after 44 years of regular occupation, the Spanish garrison was sent to Montevideo because of the War of Independence that had started in Buenos Aires in 1810 This war led to the Declaration of Independence in 1816, following the North American example

Argentina, as heir to the Spanish possessions, took charge of the Islas Malvinas in 1820 and left a representative of the Government of Buenos Aires

In 1825 Britain recognized the independence of Argentina and signed a treaty of peace and commerce without questioning the Argentine possession over the islands

In 1829 the administration of the island was reorganized and a governor was sent from Buenos Aires He settled there with his family

On 2 January 1833 John James Onslow, commanding the British frigate "Clio", dismissed the Argentine authorities by force The British forces stayed there until 2 April 1982

The claims: The Argentine government considered the British action to be plunder and immediately made the corresponding claim Since then this claim has been repeated by the Argentine government (regardless of its political affiliation) at every convenient opportunity but no satisfactory answer has been obtained

The creation of the United Nations opened a new possibility, particularly when in 1960 it decided to end colonialism Britain spontaneously presented the "Falkland Islands" as a "non-autonomous territory", an expression that replaced the old term "colony" (Resolution No 1514, XV General Assembly)

Since Argentina claimed the property of the islands, in 1965, after patient diplomatic procedures, the 20th General Assembly issued Resolution No 2065 (approved by 94 votes in favour, none against and 14 abstentions, Britain among them) in which both Britain and Argentina were invited to solve the conflict in a peaceful way

Conversations started and during this time Argentina has steadily contributed to the improvement of the living conditions of the islanders which were neglected by the British government Schools were improved, an airport was built (the only one in the islands) and twice-weekly air communication with the continent through an Argentine airline was started, thus ending the island's long isolation Free health assistance and education on the continent were also offered to the islanders It is important to stress that the islanders were primarily dependent on the "Falkland Islands Company" which controlled every aspect of their lives production, commerce and even food supply The company also owns most of the useful land

At first Britain seemed to accept a diplomatic solution and even stated that she would transfer the islands to Argentina provided the interest of the 1,800 inhabitants was taken into consideration However, Britain subsequently paid no further attention to the subject and no more steps were made in the negotiation

The conflict: The present conflict originated as follows At Grytviken, San Pedro Island, a Malvinas Islands dependence, there was on old and abandoned whaling factory An Argentine scrap dealer made the proper arrangements with the owner of the factory to dismantle it and sell it for scrap The arrangements were made according to British law and with the British authorities' approval Nevertheless, once the Argentine workers were at Grytviken, they were considered to be "invaders" and a warship was sent to clear them out Argentina in turn first sent another warship to protect the workers and then occupied the islands The recovery of the whole archipelago was made without any harm to the British Not a single drop of British blood was shed but four Argentine people lost their lives

ALBERTO C TAQUINI
Asociacion Argentina Para el Progreso de las Ciencias, Buenos Aires, Argentina

Logarithmic SI

SIR — I have for some time made private use of a notation for expressing numbers and measurements that others might perhaps also find useful It was suggested to me by the symbol "*p*H"

The first step is to let the prefix p mean "antilog to the base 10" So p0 is 1, p6 is 1,000,000, 2,000 is 2p3 or, to one decimal place, p3 3 For, say, 0 02 there is a choice of $\bar{p}$2 3, $\bar{p}$1 7 or 2$\bar{p}$2 (More generally, p may be regarded as p^1, giving $p^0 2$ as 2, $p^1 2$ as 100, $p^2 2$ as p100, and so on — so $p^9 9$ is a very large number)

The next step is to let pQ0 mean the SI unit value of some quantity Q, this generates a notation that is particularly useful for very large or very small measurements

My own practice is to use pL0, pM0 and pT0 for 1m, 1kg and 1s respectively Area, volume, speed and acceleration may conveniently be written pL^2, pL^3, pL* and pL** (I say 'pL stars' — c f x), the two latter pL/T and pL/T^2 alternatively For other units I use the SI symbol itself — 1 joule is pJ0 (I also write pD0 for 1 kg m^{-2})

As examples, electron mass and radius (m_e and r_e) are $\bar{p}$M30 0 and $\bar{p}$L14 6, the solar mass pM30 3 and the Solar System's diameter about pL13, the speed of light $\bar{p}$L*8 5 and of continental drift about pL*9, and a year is within 1 part in 500 of pT7½

This notation is particularly practical in that it avoids the awkwardness of the half-sized superscript (2×10^{30} kg), while getting round the somewhat clumsy SI prefix system, which in general usage has caught on only rather patchily

J A NICOLL
London W5, UK

Support the Zoo

SIR — The article by Jane Wynn in *Nature* 13 May (p 97) points up what seems to me to be an obvious anomaly Kew Gardens is (rightly) given enough Exchequer support to enable it to charge an entry fee of 10p, and to have over 1 million visitors a year London Zoo, which has similar functions, is forced to try to charge an "economic" entrance fee and hence price itself out of existence Why the difference?

J R BAKER
The Institute of Terrestrial Ecology, Cambridge, UK

NEWS AND VIEWS

The oncogenic circle closes

from Peter W.J Rigby

A LITTLE over a year ago I discussed in these columns[1] recent work from the laboratories of Geoffrey Cooper (Harvard Medical School) and Robert Weinberg (MIT) which showed that the NIH3T3 line of mouse fibroblasts can be morphologically transformed by transfection with DNA from chemically induced and spontaneous tumour cell lines The use of DNA-mediated gene transfer techniques to define and characterize the oncogenes of non-virally induced tumours marked a significant step forwards in molecular oncology Yet in March 1981 many questions remained Nothing was known of the nature of these oncogenes or of their products and there was much speculation as to the relationship between these cellular transforming genes and proto-oncogenes, the cellular homologues of the oncogenes carried by strongly transforming retroviruses[2] Now an extraordinary series of reports from the groups of Cooper and Weinberg and of Michael Wigler (Cold Spring Harbor Laboratory), Mariano Barbacid, Edward Scolnick and Douglas Lowy (National Institutes of Health) has precisely identified several cellular oncogenes and their products and has shown that in two cases, the genes detected by transfection are proto-oncogenes

Cellular oncogenes were first characterized in terms of their sensitivity to restriction enzymes It was shown that the DNA of different tumour types transferred different oncogenes but that within a tumour type, for example, mammary carcinomas, the same oncogene was transferred by the DNAs of independent tumours even when the tumours arose in different species, suggesting that the oncogenes are specific for a particular cell type This idea has now been extended by Cooper's group in a study of twenty B- and T-lymphocyte neoplasms of human and murine origins[3] that shows that different oncogenes are activated in tumours derived from cells of the same lineage at different stages of differentiation Either different oncogenes are highly susceptible to activation at different stages of differentiation or the expression of a particular oncogene is required to transform cells of a given differentiation state This latter explanation, of target specificity within a lineage, has parallels in the interactions of defective avian leukaemia viruses with their target haematopoietic cells[4]

The expected application of recombinant DNA techniques has now occurred, the favourite target being the human bladder carcinoma oncogene which has been cloned by the groups of Barbacid, Weinberg and Wigler[5–7] The gene is contained within an approximately 6 6 kilobase (kb) *BamI* fragment and Barbacid's group has further localized the transforming activity to a 3 0 kb *SacI* fragment[8] Southern blotting experiments show that NIH cells transformed by bladder carcinoma DNA always carry this fragment of human DNA and that the allele of the gene carried in bladder carcinoma cells is not detectably rearranged in comparison with the allele present in normal human DNA Barbacid's group has cloned the homologous sequence from normal human fetal liver DNA and shown that within the biologically active 3 0 kb *SacI* fragment there are no differences detectable by restriction enzyme or heteroduplex mapping[8] The fact that the T24, J82 and EJ lines of bladder carcinoma cells transfer the same oncogene is not surprising because at a recent meeting in Squaw Valley, California*, Wigler reported restriction site polymorphism and isoenzyme data which show that the T24 and J82 bladder carcinoma lines are identical and that the EJ line is, at the least, very closely related At the same meeting, Cooper reported the cloning of another oncogene transferred by DNA from bursal lymphomas of the chicken The striking aspect of this gene is its small size, only 1 8 kb of DNA are required for biological activity and much of this segment is comprised of highly repeated sequences The region that hybridizes to mRNA is very short, approximately 0 3 kb, suggesting that the gene product will be a rather small polypeptide

The relationship between cellular oncogenes and proto-oncogenes has been studied by all the groups About fifteen proto-oncogenes are now known and they have acquired a genetic notation of their own[9] Each gene has a three letter symbol derived from the name of the virus in which it was first defined Thus *abl* is the transforming gene of the Abelson murine leukaemia virus, *fes* the gene of feline sarcoma virus, *myc* the gene of avian myelocytomatosis virus (MC29) and so on The prefix v- indicates the intronless version of the gene carried by the retrovirus, c- the homologue in normal cellular DNA

Cooper and Weinberg and their colleagues have used cloned retrovirus *onc* gene probes in Southern blotting experiments to analyse the DNAs of NIH cells transformed by a variety of cellular oncogenes[10,11] The mouse homologues of the *onc* genes are always detected but if the donor DNA is of a different species, say human, and if the transferred cellular oncogene is related to the retrovirus probe, the question remains whether additional fragments corresponding to the transferred human gene will be detected To the great delight, and probably even greater surprise, of all retrovirologists, such relationships have been revealed in the first series of experiments Weinberg's group has shown that the EJ bladder carcinoma oncogene is homologous to c-Ha-*ras1*, the cellular equivalent of the transforming gene of Harvey murine sarcoma virus (HaMSV) Cooper's group report similar data for the EJ and J82 oncogenes but even more surprisingly show that the oncogene transferred by the DNA of the LX-1 human lung carcinoma line is homologous to v-Ki-*ras*, the transforming gene of the Kirsten murine sarcoma virus (KiMSV) Wigler (Squaw Valley meeting) has also reported the same results Weinberg's data show that a novel fragment hybridizing to c-Ha-*ras1* is present in all NIH cells transfected with EJ or T24 DNA and in secondary transfectants Moreover, if DNA from a secondary EJ transfectant is cleaved with *BamI*, an enzyme known not to inactivate this oncogene, and transfected into NIH cells, one again sees a novel fragment homologous to c-Ha-*ras1* These novel fragments are also detected by the cloned oncogene and both probes hybridize to the same fragment of normal human DNA The groups of Barbacid and Weinberg have compared the cloned *ras*

Peter W J Rigby is a member of the Cancer Research Campaign Eukaryotic Molecular Genetics Research Group and a Senior Lecturer in Biochemistry at the Imperial College of Science and Technology, London SW7 2BZ

*The Cetus UCLA Symposium on 'Tumor Viruses and Differentiation', was held at Squaw Valley, California, 21–27 March 1982

0028-0836/82/230451 03$01 00

gene and the cloned bladder oncogene[8,11], and have shown that the sequence homology resides within the 3 0 kb *SacI* fragment required for biological activity Quite unexpectedly, Barbacid's group also demonstrates weak homology between the T24 oncogene clone and v-*myc*, although this homology lies within a 0 85 kb *SacI* fragment not required for biological activity

These data show clearly that the 3 0 kb *SacI* fragment contains the bladder oncogene and that there is homology to *ras*, but do not prove that the two genes are identical, they could merely be closely linked or interdigitated However, analyses of expression make this latter possibility highly unlikely Wigler's group showed[7] that the T24 oncogene detects an approximately 1 2 kb cytoplasmic RNA in T24 cells and in NIH cells transformed by the T24 oncogene but not in untransformed NIH cells Weinberg's group has now shown[11] that this RNA, and a 5 1 kb transcript, hybridize to both the cloned oncogene and c-Ha-*ras*1, and Barbacid's group reach the same conclusion[8] using the cloned oncogene and v-*bas*, the transforming gene of the BALB murine sarcoma virus, which is identical to *ras* Final evidence for identity comes from studies at the protein level Scolnick's group has extensively analysed the protein encoded by v-*ras* and has raised a panel of monoclonal antibodies against this phosphoprotein of apparent molecular weight 21,000 (p21) Barbacid's and Cooper's groups show[8,10] that NIH cells transformed by the bladder oncogene contain higher levels of p21 than do untransformed cells and Weinberg's group mention that they have similar data[11]

Thus the cloned human bladder carcinoma oncogene is homologous to *ras*, the transforming gene of HaMSV, cells transformed by the bladder oncogene express an RNA which hybridizes to both oncogene and *ras*, and such transformed cells contain elevated levels of the protein encoded by *ras* The inescapable conclusion is that the bladder oncogene is an allele of the normal human homologue of *ras*, activated in a subtle, yet to be determined way If this is the case, then the normal human gene, if artificially activated, should transform NIH cells Lowy and Scolnick and their colleagues have shown that this does indeed occur As part of their wide-ranging studies of the *ras* system they have cloned four segments of human DNA with homology to *ras*, one of which, human c-Ha-*ras*1, is very similar to the rat c-Ha-*ras*1 probe used by Weinberg This cloned segment does not transform NIH cells If, however, the 3 0 kb *SacI* fragment from this clone, which is, of course, equivalent to the 3 0 kb *SacI* fragment of the bladder oncogene, is linked to the long terminal repeat (LTR) of HaMSV, the resultant chimaeric molecule transforms Similar activation can be achieved using the LTR of feline sarcoma virus The transformed cells contain within their DNA the human c-Ha-*ras*1 gene linked to the viral LTR and express the p21 protein

A normal human gene can thus be activated either by some as yet undefined change occurring during the induction or progression of a human tumour or by linkage to the control sequences contained within a retroviral LTR The consequence of this activation is the expression of elevated levels of p21 which is also the transforming agent of a murine retrovirus It should be noted that transformation does not result simply from the expression of p21 but from elevated expression of this protein Both the corresponding RNA and p21 are found at low levels in normal human cells (ref 12 and Wigler at the Squaw Valley meeting) The notion that cancer is a disease caused by dosage, that it results from elevated expression of normal gene products, is thus further supported

Cooper and Weinberg have also taken a separate approach to the identification of the products of cellular oncogenes[13,14] NIH cells transformed by a cellular oncogene can be used to induce tumours in young mice and sera from the tumour-bearing animals can then be employed to immunoprecipitate labelled extracts from cells transformed by that oncogene Any proteins specifically immunoprecipitated are likely to be either the product of the oncogene or a protein directly induced by its action Sera taken from animals bearing tumours induced by NIH cells transformed by other oncogenes should not recognize this protein and, conversely, the protein(s) should not be present in extracts of cells transformed by other oncogenes In this way, Weinberg's group has identified a non-glycosylated phosphoprotein of apparent molecular weight 185,000 synthesized in NIH cells transformed by the oncogene of nitrosoethylurea-induced rat neuroblastomas This protein is also found in cells transformed by DNA from a rat glioblastoma but this tumour was obtained by the same protocol as the neuroblastomas and derives from a similar tissue type Cooper's group has concentrated on cells transformed by the oncogene of the human mammary carcinoma line MCF-7 The antisera obtained immunoprecipitate a non-phosphorylated glycoprotein of apparent molecular weight 86,000 (gp86) and often a protein of apparent molecular weight 19,000 Restriction enzyme digestion experiments have shown that the MCF-7 oncogene is closely related to the oncogene of mouse mammary carcinomas induced by mouse mammary tumour virus (MMTV) or by dimethylbenzanthracene (DMBA) In agreement with this result, gp86 is precipitated from extracts of cells transformed by the MCF-7 oncogene by sera from mice bearing tumours induced by cells transformed by DNA from MMTV- and DMBA-induced mammary carcinomas and by sera from mice bearing primary MMTV-induced carcinomas gp86 cannot be detected, however, in cells transformed by DNA from MMTV- or DMBA-induced carcinomas, or in primary MMTV-induced carcinomas or in MCF-7 cells Thus while the connection between gp86 and the cellular oncogene is not clear-cut, it seems likely that this protein will turn out to be either the product of the mammary carcinoma oncogene or to be directly induced by it

Despite the astonishing advances that have been made in the past year, many important questions remain unanswered First, what is the general validity of the NIH3T3 cell assay for cellular oncogenes? Almost all the work has used this cell line simply because it transfects extremely efficiently These cells are not normal however, and could be described as teetering on the edge of transformation Many are the tales of those who, in their first experiments with these cells, succeed in generating many foci without the application of any DNA! In more scientific terms, these cells transform with kinetics indicating that only a single event is required, whereas most data suggest that natural carcinogenesis is a multi-step phenomenon However, Sager and her colleagues have recently shown that the bladder carcinoma oncogene will efficiently transform CHEF/18 cells[15] This Chinese hamster embryo fibroblast line has been extensively studied by Sager's group and their data indicate that transformation in this system normally requires multiple events They conclude that the activated bladder oncogene is the end product of the multi-step carcinogenic process Weinberg's group has shown[14] that the rat neuroblastoma oncogene will transform Rat-1 cells, a line of rat fibroblasts that (in my experience at least) is more normal than NIH3T3 cells The development of additional cell systems for the assay of oncogenes will be important but in the final analysis we need to know whether an oncogene will transform its apparent target cell type However, the development of efficiently transfectable cells derived from a wide variety of specialized tissues will be technically difficult

The second major question is whether

1 Rigby, P W J *Nature* **290**, 186 (1981)
2 Bishop, J M *Cell* **23**, 5 (1981)
3 Lane, M-A *et al* *Cell* **28**, 873 (1982)
4 Hayman, M J *J gen Virol* **52**, 1 (1981)
5 Pulciani, S *et al* *Proc natn Acad Sci U S A* **79**, 2845 (1982)
6 Shih, C & Weinberg R A *Cell* (in the press)
7 Goldfarb, M *et al* *Nature* **296**, 404 (1982)
8 Santos, E *et al* *Nature* (in the press)
9 Coffin, J M *et al* *J Virol* **40** 953 (1981)
10 Der, C J *et al* *Proc natn Acad Sci U S A* **79**, 3637 (1982)
11 Parada, L F *et al* *Nature*, this issue, p 474
12 Chang, E H *et al* *Nature*, this issue, p 479
13 Becker, D *et al* *Proc natn Acad Sci U S A* **79**, 3315 (1982)
14 Padhy, L C *et al* *Cell* **28**, 865 (1982)
15 Smith, B L *et al* *Proc natn Acad Sci U S A* **79**, 1964 (1982)
16 Hayward, W S *et al* *Nature* **290**, 475 (1981)
17 Payne, G S *et al* *Cell* **23**, 311 (1981)
18 Neel, B G *et al* *Cell* **23**, 323 (1981)
19 Cooper, G M & Neiman, P E *Nature* **292**, 857 (1981)

oncogenes detected by the NIH cell assay are the primary causative agent of the tumour from which they are derived or is their activation the end result of a complex series of genetic events? That this latter possibility may be correct is suggested by work on bursal lymphomas of chickens If chickens are infected with lymphoid leukosis viruses, avian retroviruses which do not carry an *onc* gene, bursal lymphomas arise with long latency periods Analysis of the DNA of such tumours shows[16-18] that the retrovirus integrates so that its LTR, the segment of the viral genome containing both a promoter and an enhancer, is adjacent to c-*myc* and transcription of this proto-oncogene is thus activated However, if bursal lymphoma DNA is used to transform NIH3T3 cells the oncogene transferred is not c-*myc*[19] but the gene now cloned by Cooper and discussed above By analogy with this situation the primary target during the induction of bladder carcinoma may not be the gene which has been cloned If so, quite different experimental approaches will be needed to identify it, and the oncogenes of the many tumours not active in the NIH cell assay

There can, however, be no doubt that the work discussed here is a very significant landmark in the development of our understanding of carcinogenesis The application of modern genetic and immunological techniques has led to insights undreamed of even a few years ago The argument that fundamental research in molecular biology is irrelevant to the cancer problem is surely truly dead and buried □

In, out and about meteorites

from N J McNaughton and P K Swart

AT a recent conference* in Houston much attention was given to the search for evidence of what must have been (if, indeed, it did occur) the most dramatic interaction between the Earth and an extraterrestrial object — the impact of the 10 km diameter meteorite that Alvarez has proposed was responsible for the mass extinction of the dinosaurs and many diverse faunal and floral groups at the end of the Cretaceous period Alvarez based his hypothesis on the observation of anomalous iridium levels at the Cretaceous–Tertiary (K–T) boundary (*Science* **208**, 1095, 1982) and, judging from the conference, a growing number of scientists are now committed to his view

Theoretical and experimental studies clearly demonstrate both the catastrophic consequences of a meteorite impact in the ocean and the global dispersal of ejecta A massive 5 km high tidal wave (Ahrens and O'Keefe, Caltech) would be produced and disrupt food chains by extensive stripping of sediments and silting A worldwide dust layer would be raised by the impact Atmospheric temperatures would fluctuate rapidly as a result of the production of volatilized steam and ejecta from the impact and then drop as ejected dust obscured the Sun To understand how the possible extinction mechanisms operated, scientists have studied the isotopic composition of the ejecta, particularly the layers of sanidine, glass spherules and microtektites which coincide with the extinct marine phytoplankton in the deep-ocean sediments

Epstein (Caltech) showed that the sanidine spherules have a terrestrial ratio of ^{17}O to ^{18}O, but their unusually heavy $\delta^{18}O$ may be due to either exchange with atmospheric oxygen at the very high temperatures that would accompany impact, or diagenesis DePaulo's group (University of California, Los Angeles) also noted the high $\delta^{18}O$ of the K–T boundary and further suggested from Sr and Nd isotopic studies that the layer resulted from an oceanic impact in which most of the material is of terrestrial origin, but only a small proportion is in fact derived from the local sea floor Shaw and Wasserburg (Caltech) argue from their Sr and Nd isotopic data that the sanidine spherules are of non-oceanic derivation Although this apparent conflict undoubtedly arises from the different types of samples studied, all the isotopic evidence supports the view of Ahrens and O'Keefe that the meteoritic component in the ejecta is minor

The impact site has not been identified as neither a continental nor an oceanic crater of the appropriate size and age has been found An oceanic impact would seem more likely, however, because a greater proportion of the Earth was ocean at that time Over half the late Cretaceous oceanic crust has now disappeared into subduction zones, and the lack of evidence of an oceanic crater is thus not necessarily a problem for the hypothesis

While the ejecta of meteorite impacts on Earth are sought for further study, the debris from similar collisions on other planets may already be represented in our meteorite collections The achondritic shergottites, nakhlites and chassignites (SNC), represented by only nine of the several thousand meteorites held in museum collections, are argued to have an origin in meteorite impact with asteroid or planet The SNC are distinguished by relatively young ages, 1 3 billion years, unique but consistent oxygen isotope composition (Clayton and Mayeda, University of Chicago), unusual 'cumulated' texture indicative of crystal settling under gravity, and complex chemical composition The most obvious initial source, the asteroids, are precluded as these have insufficient mass to sustain the igneous activity that could account for the young age Sufficient melting could occur, however, as a result of collision between the meteorite and either an asteroid or a planet such as Mars

Opponents of the idea that SNC meteorites are martian ejecta argue that the proposed 100 m diameter SNC parent body could not reach the required escape velocity of 5 km s^{-1} An impact of sufficient size would vaporize, melt and crush to fine dust most of the ejecta (Singer and Melosh, University of New York) and, in any case, the resultant fragment should be highly shocked, but this is not observed Singer and Melosh propose instead that SNC meteorites repesent 'pockets' of melted rock caused by impacts on large asteroids The proponents of the martian impact theory counter that such melt sheets cannot produce the required textures and point to chemical similarities between martian soil and SNC meteorites Nyquist (Johnson Space Center) argues that the problem of reaching the escape velocity may be overcome if incoming projectiles were to strike the surface of Mars at a shallow angle, so that debris ricocheted away at high velocity Such impacts would leave characteristic 'butterfly' pattern craters which are, indeed, plentiful on Mars A more rigorous chemical and isotopic comparison with martian data obtained from the Viking programme or, in the future, with returned samples, will perhaps resolve the origin of the SNCs

To understand chondrule formation and meteorite agglomeration, meteoriticists are increasingly turning to the stable isotopes of the light elements for new insight Clayton and Mayeda (University of Chicago) demonstrated that two separate processes are recorded in the oxygen isotope composition of density separates from Murchison The heavy-density separates, ostensibly from the chondrules, and olivine-pyroxene and spinel separates define a line of slope around 1 on a plot of $\delta^{17}O$ versus $\delta^{18}O$ which is best explained by high-temperature mixing and exchange between ^{16}O-rich solids with ^{16}O-poor gas The low-density and calcite separates define a line of slope around 0 5 which diverges from the mixing line and is explained by low-temperature alteration of the pre-existing materials by water from the same gaseous reservoir that produced the mixing line That the carbonate, in particular, is formed significantly later

*The 13th Lunar and Planetary Science conference was held in Houston, Texas on 15–19 March 1982

N J McNaughton and P K Swart are at the Planetary Sciences Unit, Department of Earth Sciences, University of Cambridge, Cambridge CB2 3EQ

0028-0836/82/230453-02$01 00

than the chondrules and the high-temperature silicates is further supported by new Rb/Sr data for Orgueil, which suggest an age gap as large as 160 Myr (Macdougall and co-workers, University of California, San Diego) The secondary nature of the carbonates has important implications, not the least of which is its enrichment in carbon-13 compared with other co-existing carbon-bearing phases

Several workers have attempted to separate the carbonaceous components by HF/HCl demineralization of the meteorites Kerridge (University of California, Los Angeles) found that stepwise oxidation of the acid residue from the CM2 chondrite Murray yielded a minor phase which did not oxidize until over 1,000°C and had a $\delta^{13}C$ and $\delta^{15}N$ of around $+100°/_{oo}$ (PDB) and $+90°/_{oo}$ (AIR) respectively, as well as a low δD A carbonaceous phase of this nature has not been recognized either from other meteorites or from Murray by workers in the field However, attempts to confirm Kerridge's data should renew interest in identifying and explaining the phase Acid-insoluble residues, which contain major D enrichments for some carbonaceous chondrites, also appear to be the D carrier in the primitive ordinary chondrites (Yang and Epstein, Caltech) The Cambridge–East Kilbride group also demonstrated that most type 3 ordinary chondrites have enrichments in D above the terrestrial range, but only the least metamorphosed show extreme enrichments Furthermore, the D carrier in one of the least altered type 3 chondrites, Semarkona, seems to be a minor phase, which decomposes when heated to 550–650°C and has a minimum δD of over $+5,500°/_{oo}$ (SMOW)

A controversial issue is whether ^{26}Mg excesses in Al-rich minerals are a result of *in situ* decay of ^{26}Al or whether ^{26}Mg is a relict species that was already present at the time of Solar System formation The idea of 'fossil' ^{26}Mg was again forcefully argued by Clayton (Rice University) as being the more plausible explanation because of mixing constraints in the early solar nebula, and because of certain other isotopic anomalies such as ^{16}O and ^{50}Ti Advocates of the *in situ* decay model point to data obtained using the ion microprobe (Huneke, Caltech), which show uniform distribution of ^{26}Mg in minerals from Allende □

Hard facts on structure: hot air about mobility

from David M Blow

THE idea that the inherent flexibility of biomolecules relates to their function received another airing at the CIBA Foundation Meeting* on March 2-4, 1982 Improvements of technique now make available a wide portfolio of techniques for the measurement of mobility in macromolecules The most important is NMR, which allows separate relaxation times to be measured for every line in the spectrum Almost any other spectroscopic technique, from fluorescence to EPR, can give highly specific information about the mobility of a chromophore or a spin label Time-resolved X-ray diffraction, which in principle can give a 'flash' picture of a structure any time after triggering an event, is still in its infancy But X-ray crystallography can measure the time-averaged motion of individual atoms, so long as the motion is not too large, and can delineate domains of correlated motion

Computational capacity has increased to the point where all the atomic interactions for a small protein can be simulated, allowing the dynamics of the molecule to be calculated over a period of a few picoseconds The achievable time span is far too short to cover a significant chemical event, but the results do appear realistic so far as they go Martin Karplus (Harvard University) has used simulations to measure correlation times of a few picoseconds between atomic motions, and has estimated the apparent viscosity of the protein fluid in which an individual atom moves People speak of analysing the normal modes of vibration of a protein molecule, though nobody has done so yet

What has all this to do with the biological function of a molecule? Fred Richards (Yale University), stern but perceptive chairman of this rather unruly symposium, stressed this question in his introduction Is mobility essential or incidental to biological action? What experiments and theories can be devised to establish a relation between them? He got few answers

The most thoughtful came from Greg Petsko (MIT) who stressed that in many cases, access to an active site, and release of product, would be greatly hindered, if not prevented, if the adjacent structure were rigid He cited triose phosphate isomerase and pancreatic ribonuclease as examples where substrate binding is accompanied by the ordering of an adjacent mobile part of the structure — a kind of flexible lid which closes over the substrate The entropy change of such a process destabilizes the enzyme–substrate complex and thus facilitates release of products In cases where a linked function is invoked, this ordering of a mobile chain may convey a signal to a part of the molecule which the disordered chain can touch In the pyruvate and 2-oxoglutarate multi-enzyme complex, Gordon Roberts (National Institute of Medical Research) and Richard Perham (University of Cambridge) showed that lipoyl-lysine residues act like flexible arms to allow the transfer of substrate from the active site of one enzyme subunit to another

Many examples of mobile molecules and many different techniques were discussed Nucleosomes show a high degree of mobility on the nanosecond to microsecond time scale, making them a good subject for NMR study Michael Hogan (Princeton University) has used ^{31}P magnetic resonance to study the motions of DNA in both the isolated state and in nucleosomes, supplementing these studies with fluorescence anisotropy decay measurements using DNA-binding dyes The observed motion suggests twist or flexure of the DNA helix Mort Bradbury (University of California, Davis) reported NMR mobility studies on the histone and non-histone proteins of nucleosomes The unusually high mobility of these proteins seems to result from their highly charged character, which causes them to be fully or partly unfolded under almost all conditions

Bob Williams (University of Oxford) suggested that the action of calcium-binding trigger proteins depended on motion of hydrophobic groups detectable by NMR In muscle, motion of the 'head' part of the myosin molecule, which forms a cross-bridge linking the actin and myosin filaments, achieves relative motion between them Measurements reported by Stefan Highsmith and Oleg Jardetzky (University of California, San Francisco) showed that mobility in parts of the myosin molecule was quenched on binding to F- or G-actin David Thomas (University of Minneapolis) had studied the same system by EPR and spin label, and found a two-state 'relaxed' and 'attached' system with different degrees of order, modulated by calcium The function of muscle clearly depends on molecular motion, but are these motions related to mobility on the NMR and EPR time scales? Bill Harrington (University of Baltimore) went so far as to suggest that the rotation of the myosin head was achieved by a transition between a folded and an unfolded state in the stem of heavy meromyosin which carries it

Any dissatisfaction felt about these inconclusive speculations was dispelled by two exciting PMR papers which gave clear

*The proceedings of the meeting will be published by Pitman Books, London, in winter 1982 under the title *Mobility and function in proteins and Nucleic acids* (Ciba Foundation symposium 93)

David M Blow is Professor of Biophysics and Dean, Royal College of Science at Imperial College, London SW7 2BZ

0028-0836/82/230454-02$01 00

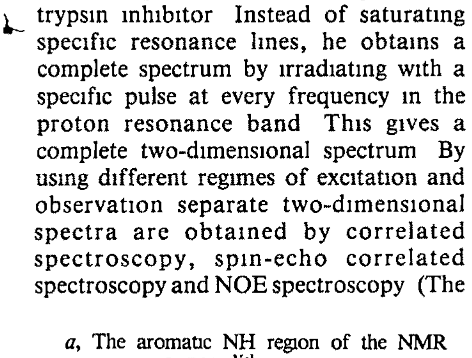
and definite results, using the nuclear Overhauser effect (NOE) to detect dipolar interactions between protons which are close together If the relative populations of the nuclear spin energy levels of one proton are perturbed — for example, by saturating its resonance frequency — the dipolar interactions transmit the perturbation to adjacent protons, resulting in an altered resonance intensity Pulses may be specifically shaped to avoid exciting the H_2O resonance Redfield has pioneered the use of this technique in the analysis of tRNA spectra, and has been able to assign resonance lines in the D stem of $tRNA^{Asp}$, and also in reverse-Hoogsteen base pairs (*Nucleic Acids Res* **9**, 7073, 1981)

Brian Reid (University of Seattle) has applied this technique to the aromatic NH band, which contains just one resonance for each base pair When such a resonance is perturbed, either one or two other NH resonances show up by NOE These are resonances in adjacent stacked base pairs An example is shown in the figure In this way Reid has been able to survey all the stacked base pairs, and to assign all the corresponding 28 ring NH resonances, in $tRNA^{Tyr}$

Kurt Wuthrich (University of Zurich) is tackling with Richard Ernst the more ambitious problem of the basic pancreatic trypsin inhibitor Instead of saturating specific resonance lines, he obtains a complete spectrum by irradiating with a specific pulse at every frequency in the proton resonance band This gives a complete two-dimensional spectrum By using different regimes of excitation and observation separate two-dimensional spectra are obtained by correlated spectroscopy, spin-echo correlated spectroscopy and NOE spectroscopy (The engaging acronyms COSY and SECSY were coined for the first two, the last having a more nasal connotation) COSY has enabled NH and $C^{\alpha}H$ resonances in the same amino acid to be correlated, while NOESY allows the identification of β-structure interactions by the proximity of $C^{\alpha}H$ and NH protons in adjacent residues (*J molec Biol* **155**, 311, 1982) By further development of these methods, almost every resonance in glucagon and in the pancreatic trypsin inhibitor has been assigned Individual changes of internal mobility can be observed as pressure and temperature are changed

The claim that a three-dimensional protein structure may, in the near future, be totally determined from NMR data is beginning to look more realistic, if only for a very small, rigid protein molecule □

a, The aromatic NH region of the NMR spectrum of $tRNA^{Val}$ (*E coli*) The imino NH proton resonances lie between -9 and -15 p p m In this region only the resonances whose exchange is slowed down by hydrogen bonding appear Each represents an NH in one hydrogen bond of a base pair *b*, Saturation of the resonance at -14 4 p p m leads, by the NOE, to a signal at -7 6 p p m from a non-hydrogen-bonded resonance in the same base pair, and also to two signals at -11 7 and -12 5 p p m from adjacent hydrogen-bonded imino groups These two signals come from base pairs stacked on either side of the base pair whose resonance is saturated (Courtesy of B R Reid, University of Washington)



Do stepping stones guide axon growth?

from Hilary J Anderson

In a recent issue of *Nature* (June 3, p 404), Ho and Goodman present data supporting the idea that peripheral pioneer neurones in the grasshopper are guided to their destinations by a series of cellular 'stepping stones' The 'pioneer' neurones are of vital significance because they establish the basic pathways and tracts within the nervous system that act as guides for neurones that develop later The term was first used by Ross Harrison, in his now classic studies of frog embryos, to describe the neurones that grow out and establish connections very early in development when the distance to their target regions is very small Even though the animal subsequently grows and changes form and the pioneer tracts twist and lengthen they continue to provide the guides that allow other neurones to reach their appropriate destinations

Several mechanisms of pioneer axon guidance — mechanical channeling, basement membrane markers and cellular landmarks — have been suggested and to distinguish between them it is necessary to study the behaviour of growing neurones in some detail The tip of a growing axon is expanded into a growth cone from whose leading edge extend numerous filopodia which are 0 1–0 2 μm in diameter and many micrometres in length and filled with actin filaments The growth cone makes 'searching' movements over its substrate, extending and retracting its filopodia, and probably plays a fundamental part in the interactions with the environment that direct and control neurone growth Harrison proposed that the pioneers were guided by the configurations of tissues in their environment, grooves and spaces in the embryonic tissue providing "paths of predilection" along which the growth cones progress This idea has been supported by several recent studies using electron microscopy which show that developing or regenerating axons often grow through preformed channels or spaces (see *Nature, News & Views* **283**, 428, 1980) In other instances, however, pioneers grow over regions seemingly devoid of obvious physical regularities, suggesting that there are other factors that direct pioneering axon growth

Pioneering axon growth in the central and peripheral nervous systems of insect embryos has proved particularly amenable to experimental study The central pioneers that will be discussed below are those that establish the primary longitudinal tracts of each ganglion in the central nervous system (CNS) (see ref 2 and *News & Views* **293**, 510, 1981) Ho and Goodman studied the peripheral pioneers which establish the peripheral nerves linking the CNS and the appendages (see also refs 3–6) These neurones have been observed as they grow in the living embryo, and examined by electron microscopy, by staining with intracellularly injected fluorescent dyes and by staining with particular antibodies In this way several different aspects of their behaviour have been successfully described

In their recent paper Ho and Goodman establish the important point that the pioneers are not a special class of cell but may be motor neurones, interneurones, or sensory neurones Furthermore, the pioneers themselves may grow along other pioneers for short distances during growth It thus seems likely that the mechanisms underlying pioneering axon growth are common to all growing neurones

Electron microscopy has shown that a basement membrane lies over the neuro-ectodermal layer from which the CNS develops[2] The central pioneer growth cones always track along this basement membrane and do so at a consistent location and in a consistent direction to form a characteristic tract[2] Similarly, in the periphery, where the embryonic appendages are hollow epithelial buds, the pioneer growth cones grow only over the inner surface of this epithelium[3-5] and do so only at certain locations to form characteristic pathways[3-5] It has been suggested that cues within the basement membrane might provide directional information for growth cone navigation[2] Such cues would be imposed on the basement membrane by the underlying sheet of cells and could take

Hilary J Anderson is a staff scientist at the European Molecular Biology Laboratory, 6900 Heidelberg

0028-0836/82/230455-02$01 00

the form of a simple indicator of longitudinal polarity, or a continuous gradient of information or a series of discrete markers.

Other observations have suggested the cellular landmark hypothesis favoured by Ho and Goodman. Before the peripheral pioneers axons appear, glia-like cells are found in the lumen of the appendage and in contact with the inner surface ofthe epithelium. These cells are found at regular intervals along the presumptive path of the pioneers[3-5] and could provide a series of stepping stones between the tip of the appendage and its base[3]. Later the glia ensheath the pioneers to form a proper nerve bundle. Ho and Goodman observed several other conspicuously located cells which could act as landmarks for the pioneers. Furthermore, the monoclonal antibody which they used for many of their observations stains growth cone filopodia very clearly and shows that they extend for some 50 μm. Thus, at no time need an axon be navigating without being in 'filopodial grasp' of a landmark cell.

The three proposed substrates for guiding pioneer axon growth now need to be tested by experimental manipulation, such as selective ablation of putative landmark cells. Ultimately these studies must lead on the one hand to an investigations of the molecular mechanisms underlying the preferential growth of pioneers over their chosen substrates and on the other hand to a study of the principles of spatial organization which generate these substrates at their appropriate locations — the fundamental unanswered question of developmental biology. □

1. Harrison, R G *J exp Zool* **9**, 787 (1910)
2. Bate, C M & Grunewald, E B *J Embryol exp Morph* **61**, 317 (1981)
3. Bate, C M *Nature* **260**, 54 (1976)
4. Edwards, J S in *Identified Neurons and Behavior of Arthropods* (ed Hoyle, G) 483 (Plenum, New York, 1977)
5. Edwards, J S & Chen S -W *Wilhelm Roux's Arch dev Biol* **186**, 151 (1979)
6. Keshishian, H *Devl Bio* **80**, 388 (1980)

Discovery of pre-galactic lithium

from Bernard Pagel

THE reality of the hot big bang is attested by galaxy redshifts, the 2.9K microwave background and evidence for evolution in the statistical properties of distant quasars. Indeed, if grand unified theories fulfill all their promises, then such basic facts as the existence of matter (rather than antimatter) and electric charge can also be taken as consequences of the big bang. These fundamental effects were imprinted on the Universe less than 10^{-35} seconds after it was born, while the last scattering of the microwave background took place about a million years later when protons combined with electrons and space became transparent. Galaxies and stars were presumably formed later still.

In an intermediate phase, when the Universe was about one second old and its temperature around 10^{10}K, neutrinos decoupled leaving an excess of protons over neutrons. During the next two minutes or so these combined in nuclear reactions to make deuterium, helium-3, helium-4 and lithium-7, production of heavier elements being aborted by the absence of stable nuclei at mass numbers 5 and 8. The resulting pre-galactic composition of cosmic matter, consisting chiefly of hydrogen and helium-4 in proportions of about 3:1 by mass, has since been modified by cycling of some of the material through stars. This has affected material in our neighbourhood by adding a mass fraction of 1 or 2 per cent of carbon and heavier elements and a similar amount of additional helium-4. Deuterium, on the other hand, is totally destroyed in matter cycled through stars, and helium-3 and lithium-7 can be both created and destroyed, so that the net effect of stellar evolution on these two species is rather uncertain.

A remarkable claim by Monique and Francois Spite, of the Meudon Observaory near Paris, that something at least very close to the pregalactic lithium abundance can be measured directly (see this issue of *Nature*, p 483), may now provide additional constraints on this view of the early Universe.

The pregalactic composition of cosmic matter depends on a number of parameters of cosmology and particle physics, notably the ratio of baryons to photons (or, equivalently, the present mean density of ordinary baryonic matter) and the number of lepton flavours, together with the net charge number and various lepton numbers that are believed to be zero or negligibly small. As a result, the determination of pregalactic abundance provides interesting constraints on these parameters and tests the consistency of the entire picture. For example, the presence of observable deuterium in the interstellar medium implies that the Universe cannot contain a sufficient density of ordinary matter to 'close' it, that is, eventually to halt the expansion by gravity, although it could be closed by non-baryonic matter such as massive neutrinos, if these exist. A still more severe upper limit to the density is provided by the observation that the mass-fraction of helium in emission-line galaxies of low oxygen abundance (in which the gas has accordingly undergone little recycling through stars) is below 25 per cent, which implies that the baryon/photon ratio (resulting from the primordial excess of matter over anti-matter) is below 4×10^{-10} or $\Omega_b < 0.03$ (assuming three neutrino flavours and a Hubble constant of 75 km s^{-1} Mpc^{-1}) where Ω_b is the fraction of the closure density supplied by ordinary baryonic matter. If, however, the density were less than about ⅓ of this upper limit, corresponding to a helium abundance of under 23 per cent, then the expected pre-galactic abundances of deuterium and helium-3 would become embarrassingly large. Thus a really accurate determination of the pregalactic helium abundance would provide a very severe test of the consistency of canonical big bang models, if only it could be carried out. Unfortunately, the prospects of finding the pregalactic helium abundance to better than 5 per cent are not very good.

What about lithium in all this? The 'cosmic' mass fraction of lithium-7, deduced from meteorites, is 5×10^{-9} and a similar amount is found in the atmospheres of the youngest stars. In older stars, however, and notably in the Sun itself, lithium is depleted because the surface layers are mixed in the course of time with

Bernard Pagel is a Deputy Chief Scientific Officer at the Royal Greenwich Observatory, Herstmonceux Castle, Hailsham, Sussex BN27 1RP, and a Visiting Professor of Astronomy at the University of Sussex

100 years ago

THE mines opened a short time since in China in the province of Chihli, with the special support and patronage of Li Hung Chang, have recently become the subject of much adventitious interest in Europe. It was believed that, with sufficient transport, one thousand tons a day could be raised for many years from the present pits, while it was said that fifty collieries of an equal size to the present one could be opened in or near Kaiping. The information, therefore, telegraphed by Reuter's agent in Shanghai that the further working of the mines had been peremptorily stopped by the Government, came with a shock to many interested in progress in China. It was stated that a censor in a memorial to the throne complained that the long galleries in the mines, and the smoke of the foreign machinery, disturbed the earth dragon, who in his turn disturbed the spirit of the Empress, who died some months ago, and who was buried about a hundred miles off. The irate spirit of the departed lady promptly took vengeance by afflicting the denizens of the palace in Peking with measles. The latter were, the censor is reported to have said, distintly traceable to the Kaiping mines, which interfered with the *fêng-shui*. The conclusion was obvious: the mines must be stopped. Such was the story told by the Tientsin correspondent of a Shanghai newspaper. The latest information from the East enables us to say that the mines are still working as usual, and there is not the slightest evidence that there is or has been any intention of interfering with them. It is even denied that such a memorial as that mentioned above has had any existence except in the imagination of a *gobemouche* at Tientsin. However this may be, it must be confessed that the petition has a Chinese ring about it, and that the method of argument is one sufficiently familiar to readers of the *Peking Gazette*.

From *Nature* **26**, 136, 8 June 1882

0028-0836/82/230456-02$01 00

deeper layers at a temperature of 2×10^6K where lithium nuclei are destroyed The degree of depletion in ordinary stars increases with age and with diminishing surface temperature because cooler stars have deeper outer convection zones A few highly evolved stars make fresh lithium, however, and yet more may be generated by cosmic-ray spallation in the interstellar medium, so extrapolation from present-day to pregalactic lithium abundance is a hazardous business

What Monique and François Spite have done is to estimate the lithium abundance of a sample of extreme subdwarfs These are old stars with extreme metal deficiency, typically a factor of 100 down on the Sun for carbon and heavier elements, so that the material from which they were formed could have undergone little previous cycling through stars Out of the five stars of differing surface temperature that they studied, the four hottest turned out to have one-tenth of the present-day lithium abundance — a strikingly large amount in view of the overall metal deficiency — and only in the coolest star was it undetectable, presumably because of depletion through convective mixing with deeper layers The authors argue that the constancy of lithium abundance in the four hotter stars indicates that none of these has been significantly affected by lithium depletion If they are right, then they have a value for the pregalactic lithium abundance, if not, then they still have an interesting lower limit

What does this new result tell us about the big bang? Unfortunately the predicted lithium abundance is a variable and generally fairly insensitive function of baryon density and, furthermore, it is rather uncertain For example, Olive and coworkers (*Astrophys J* **246**, 557, 1981) suggest an upward revision by a factor of 3 over the earlier calculations referred to by Spite and Spite, in this case the minimum possible value is minutely (25 per cent) greater than that observed, although it is striking that this value occurs at parameter values close to those deduced from helium and other light elements A fairly conservative view of the errors (especially with regard to possible depletion) suggests that lithium in subdwarfs provides an upper limit to the cosmological baryon density that is consistent with, but less stringent than that provided by helium What is exciting about the Spites' discovery is that lithium is there at all, and that its amount is not far from big bang predictions □

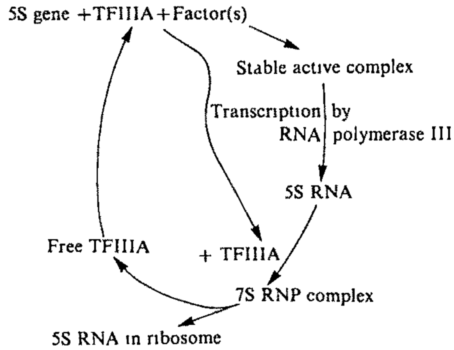
Stable gene expression *in vitro*

from H R Woodland

ONE characteristic of cell differentiation is its high degree of stability, and there is a suspicion among molecular biologists that this derives from a basic stability in the expression of genes involved in differentiation Our understanding of the issue is unlikely to improve beyond this point, however, until systems are developed which enable active genes to be reconstructed *in vitro* Bogenhagen, Warmington and Brown of the Carnegie Institution of Washington, together with R G Roeder's group at the University of Washington, may now have developed just such a system In a recent issue of *Cell*[1], they describe the construction of 5S gene protein complexes with the properties of stable gene expression or repression that have been suggested to occur in intact cells

As much is known about 5S gene transcription as for any other eukaryotic gene (see Korn's recent review in *Nature*[2]) 5S genes encode 5S RNA, a component of all ribosomes except those of animal mitochondria Like tRNA and a few other small RNAs, 5S RNA is produced by RNA polymerase III, a different enzyme from that which makes mRNA The unusual characteristics of this polymerase mean that the 5S genes differ in structure from genes encoding mRNA Surprisingly, the only part of the 5S gene actually essential for its transcription is residues 50–83 of the 120 nucleotides that are transcribed This region binds a 40,000 molecular weight protein, called TFIIIA by Roeder's group[3], which produces accurate transcription of the gene TFIIIA also binds to 5S RNA, forming a 7S ribonucleoprotein (RNP) particle This may provide a means of transcriptional regulation whereby the free protein stimulates 5S RNA synthesis, but is removed by the product of this synthesis (see the figure) A last important feature of 5S RNA, at least in *Xenopus laevis*, is that oocytes and somatic cells contain 5S RNAs of different sequence The main oocyte 5S RNA is encoded by 20,000 genes called 5S *ooc*, and that in somatic cells by a separate set of 400 genes called 5S *som*

Why do somatic cells contain different 5S RNAs? Selective transcription of the 5S *som* genes seems the cause In oocytes, some 5S *som* RNA is made, perhaps in proportion to the small number of 5S *som* genes, but the RNA seems not to enter ribosomes and is degraded[4]. Free 5S RNA is known to have a short lifetime in both oocytes and somatic cells[6-8] Unfortunately, the role of degradation in regulating 5S gene expression has never



Possible scheme of 5S RNA synthesis In a very young amphibian oocyte, 5S RNA is made in great excess of other ribosomal components and enters a 42S RNP particle, instead of a ribosome As the 42S RNP particle does not contain TFIIIA[10], the latter is released and can stimulate further 5S synthesis, just as in the normal cycle shown here

been accurately quantified in any system

The latest experiments from Brown's laboratory[1] use cell-free systems made from either oocytes or somatic cells If 5S DNAs of different types are added simultaneously to a cell-free system they compete for transcription[9] It has now been shown, however, that if two DNAs are added at different times, the first is transcribed in preference to the second[1] Indeed, if enough of the first DNA is added subsequent DNAs are not transcribed at all, no matter how much is added The first DNA seems to form a very stable, transcribable complex within a minute or two This is not an actual transcription complex, since maximal transcription is established much more slowly, thus agents other than the polymerase, but necessary for polymerase binding, must be involved in formation of the complex It might be guessed that TFIIIA is one such agent, but although 5S DNA may be bound to TFIIIA, this alone does not confer favoured transcription status upon it

Bogenhagen *et al* [1] were also able to make stable inactive 5S gene complexes Using a transcription system depleted of TFIIIA by antibody treatment, they showed that 5S DNA does not form the active complex unless further TFIIIA is added However, if histones were added instead of TFIIIA, a different complex was formed which could not then be reactivated with TFIIIA The possibility that this complex is an artefact is unlikely, since the transcription of tRNA genes is unaffected by added histones (tRNA genes do not require TFIIIA for transcription)

Bogenhagen *et al* [1] speculate that the stable complexes they have constructed *in vitro* have similar counterparts in cell-free systems This remains to be proved At least natural 5S DNA protein complexes show similar stability to those made in cell-free systems — chromatin added from somatic cells or oocytes to the same systems used to transcribe the pure 5S DNAs continues to express only those genes which were active in the intact cells from which it was derived This emphasizes that the cell-

H R Woodland is in the Department of Biological Sciences, University of Warwick, Coventry CV4 7AL

0028-0836/82/230457-02$01 00

free systems do not reprogramme chromatin, nor indeed do oocyte or somatic cell-derived systems discriminate significantly between 5S *ooc* and 5S *som* genes Why then are these genes expressed differently in oocytes and somatic cells? Bogenhagen *et al* [1] suggest that the activity of the 5S genes is determined by events during chromatin construction at DNA replication which establish an active or inactive structure that persists throughout the ensuing cell cycle The two states could be propagated through the next DNA replication unless hypothetical factors are present which change the state of the gene This hypothesis is consistent with many observations suggesting that DNA synthesis is necessary for changes in the differentiated phenotype of cells It does not fit, however, with Korn and Gurdon's[5] observation that the oocytes of certain *Xenopus* females can activate the oocyte genes of somatic nuclei in the absence of DNA replication

Bogenhagen *et al* [1] also suggest that what goes for 5S genes might also apply to genes involved in cell differentiation The dissimilarity of the transcription of the two sorts of genes by RNA polymerase III and RNA polymerase II respectively does not encourage heavy bets on this possibility Yet advances in this area certainly depend heavily on developing cell-free systems which accurately reproduce the gene regulation of intact cells There is no doubt that the new 5S systems show exciting possibilities in this direction, even if they do not yet solve the problems of cell differentiation □

1 Bogenhagen, D F , Wormington, W M & Brown, D D *Cell* **28**, 413 (1982)
2 Korn, L J *Nature* **295**, 101 (1982)
3 Honda, B M & Roeder, R G *Cell* **22**, 119 (1980)
4 Denis, H & Wegnez, M *Devl Biol* **58**, 212 (1977)
5 Korn, L J & Gurdon, J B *Nature* **289**, 461 (1981)
6 Miller J R & Melton, D A *Cell* **24**, 829 (1981)
7 Miller L *Cell* **3**, 275 (1974)
8 Perry, R P & Kelley, D E *J cell Physiol* **72**, 235 (1968)
9 Wormington, W M , Bogenhagen D F , Jordan E & Brown, D D *Cell* **24**, 809 (1981)
10 Picard B & Wegnez, M *Proc natn Acad Sci U S A* **76**, 241 (1979)

Leg 84 of the Deep Sea Drilling Project

Subduction without accretion: Middle America Trench off Guatemala

from Jean Aubouin*, Roland Von Huene†, Miriam Baltuck‡, Robert Arnott§, Jacques Bourgois*, Mark Filewicz , Keith Kvenvolden†, Barry Leinert**, Tom McDonald‡‡, Kristin McDougall‡‡, Yujiro Ogawa§§, Elliot Taylor‡‡ and Barbara Winsborough

PRESENTED below are the first results from the recent visit of the drilling vessel *Glomar Challenger* to the southern Guatemalan segment of the Middle America Trench (Leg 84 of the Deep Sea Drilling Project/International Phase of Ocean Drilling, January–February 1982) The area, shown in Fig 1, had been selected for investigation because of several interesting problems that had been posed by earlier visits to the region In 1979, Legs 66 and 67 had produced contrasting results from the two halves of the Middle America Trench divided by the Tehuantepec Ridge — on Leg 66 ocean sediment was found to be accreting against the landward slope of the trench, while on Leg 67, off Guatemala, no accretion was found Leg 67 results come mainly from one site (494) on the landward slope of the trench where basement was reached [3], drilling at three other sites were abandoned when gas hydrates were unexpectedly recovered

Seabeam [4] and deep tow instruments were also used to survey the area examined by Leg 67 The original fabric of the Cocos Plate at the East Pacific Rise is delineated by magnetic anomalies which trend 20° from the direction of the trench axis As the Cocos Plate is flexed into the trench, it breaks into oblique extensional ridges and troughs paralleling the original fabric of the Cocos Plate, and the surface expression is one of collapse into the subduction zone

On the landward slope of the trench the structural picture is less clear Palaeogene and Cretaceous rocks were recovered at its base, suggesting non-accretion off Guatemala and showing that a tectonic history exists below the Neogene slope sediments More definite conclusions could not be drawn because only one penetration had been made below the Neogene slope deposits The recovery of visible gas hydrate made further drilling unsafe because the base of the gas-hydrated sediment had, at the time, not been defined in seismic records and it was feared that free gas might be trapped at the base of an impermeable hydrated section

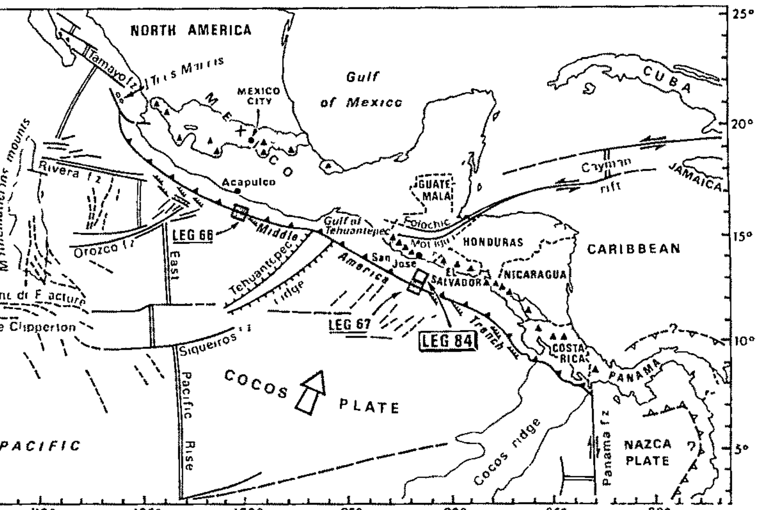


Fig 1 Sketch map of Middle America Trench

*Département de Geotectonique, Université Pierre et-Marie Curie, 4 place Jussieu, 75230 Paris Cedex 05, †Pacific-Arctic Branch of Marine Geology and ‡‡Branch of Paleontology and Stratigraphy, US Geological Survey, 345 Middlefield Road Menlo Park, California 92093, ‡Deep Sea Drilling Project A 031, Scripps Institution of Oceanography, La Jolla California 92093 (R H also at Institute of Geology, Ruhr University, Bochum, West Germany, §Department of Geology University of Oxford, Parks Road, Oxford OX1 3PR, Union Oil Company, PO Box 6176, Ventura, California 93003 **Hawaii Institute of Geophysics, University of Hawaii 2525 Correa Road, Honolulu, Hawaii 96822, ‡‡Department of Oceanography, Texas A and M University, College Station, Texas 77843, §§Department of Geology, Faculty of Sciences Kyushu University, Hakozaki Fujuoka shi 812, Japan, Espey Huston and Associates Inc , PO Box 519 Austin Texas 78767

0028-0836/82/230458 03$01 00

Leg 84 was recommended by the JOIDES advisory panels specifically to examine the tectonic and gas hydrate problems. Local base of gas hydrate reflections were found in a re-examination of predrilling site survey data and after enhancement and combination with refined temperature gradient information, the depth of the base of gas hydrate was established. Basements high above the base of gas hydrate were selected for drilling.

At the penetration sites (Fig. 2) basement was found to consist of ophiolitic rock covered by sediments of early Eocene (sites 569, 570), late Oligocene–early Miocene overlying late Cretaceous (567) and late Miocene (566) age. No age progression (younger rocks at the base to older at the top of the slope), as admitted over the supposed accreted complex off Mexico where only Neogene deposits were drilled, was found. The sites are distributed transverse to the margin and are laterally separated by as much as 100 km. Peridotite (harzburgite), gabbro, dolorite, basalt and much serpentine in various stages of decomposition were recovered. These rocks were of a metamorphic grade from low-temperature to greenschist facies and were tectonically juxtaposed without particular order. The Guatemalan margin is therefore underlain by tectonically disrupted ophiolitic rock.

The ophiolitic rock was tectonized and the tectonic complex was emplaced before the present arc-trench tectonic system. Clastics of ophiolitic rock were included in the unaltered slope sediment, particularly at the base of the slope where debris flows contain large blocks of serpentinized peridotite. Other studies[5-7] suggest the tectonic history includes three major periods:

(1) Ultramafic rock is thrust over the Jurassic–Cretaceous Nicoya complex[6] and is in turn overlain unconformably by Campanian sediments, as is evident on the Santa Elena Peninsula[8]. Initial tectonism of the Guatemalan ophiolitic rock could have started at this time, and the Cretaceous sediment at the base of the slope could correspond to the Campanian sediment cover.

(2) A strong Palaeocene uplift, documented in the Petrel well[9], may immediately pre-date the early Eocene sediment recovered at two sites.

(3) A strong Oligo–Miocene faulting left a sharp angular unconformity observed in seismic records across the edge of the shelf; the first thick hemipelagic slope mudstone deposits of the present tectonic system developed during and after this event.

There is nothing to indicate development of the present arc-trench system until the late Oligocene when the first substantial layers of volcanic ash are seen in the slope cover. The ophiolitic rock of the Guatemalan margin was certainly emplaced in pre-Eocene time and appears to be an extension of the continent of Central America. At the base of the slope

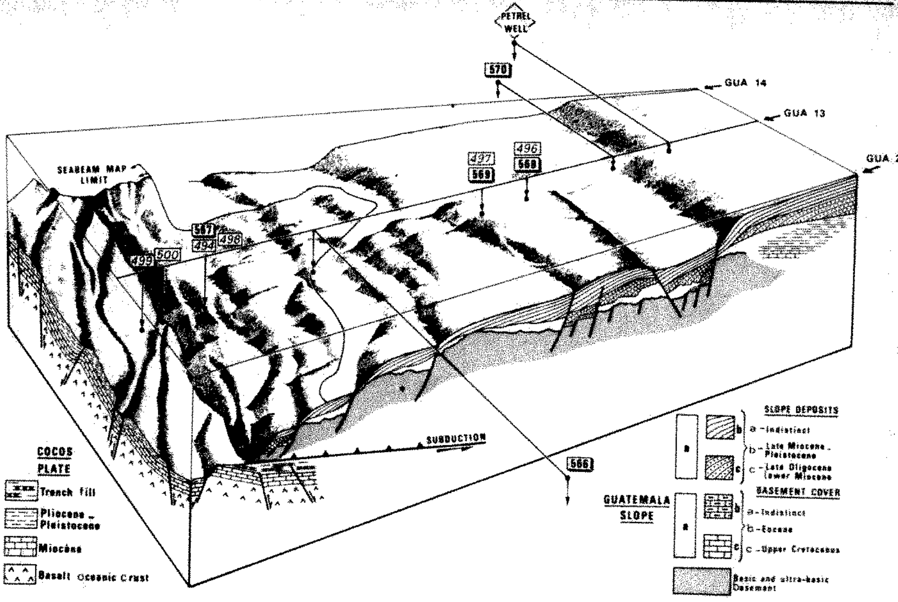


Fig. 2 Block diagram showing the convergent zone between Cocos Plate and Carribbean Plate in Leg 84 area. The morphology is reconstituted from *Jean Charcot* seabeam data and GUA 2, GUA 13 and GUA 14 seismic profiles of Texas University.

(Site 567) we came close to penetrating through the upper plate into the subduction zone. Here, Miocene to present oceanic and trench sediments are being subducted beneath the pre-Campanian ophiolitic rock at the front of the margin. Drilling ended when the hole collapsed, and the drill string became stuck. Collapse was accompanied by overpressured pore fluids — overpressure had also been observed at sites in top, mid and lower slope. Seismic records across the mid and upper slope sites show few compressional faults or folds in the slope sediment where the overpressures were observed. In fact, the most obvious structures on the transect are normal faults at the base of the slope. Elevated pore pressures are to be expected across the subduction zone, and such pore pressures help to explain the degree of decoupling required to subduct soft sediment beneath a mass of hard ophiolitic rock. High pore pressure appears to be very common in the Guatemalan convergent margin and occurs even in slope deposits that show few structures.

Many local unconformities and hiatuses were detected in sites drilled in slope deposits on the Guatemalan margin and indicate that as one depocentre is rapidly accumulating sediment, another may have lost its feeder channel and temporarily be in an interchannel or erosional area. Hollows and topographical flats from initial faulting of the surface of the ophiolitic mass are sites of ponded basins or early Miocene prograding sediment bodies. The sediment in these sequences locally contains more sand than is found in the sediment ponded in the trench axis. Sand was recovered in the channel areas, demonstrating again that the channels transport sand directly to the trench. The presence of a large amount of sand and rapid sediment accumulation is not a lithological criterion by which to distinguish between trench and slope deposits off Guatemala.

Slope deposits off Guatemala are now classic sites for gas hydrate[10] which was detected at all sites, except 567, and was recovered many times at 568, 570. The best recovery was at site 570 where a massive 3 m thick gas hydrate was cored, and samples were preserved in special pressure vessels for onshore studies. The gas hydrate was recognized: (1) visually, (2) by volumes of gas released during hydrate decomposition that greatly exceeded the volume of gas in gas-saturated water, (3) by composition — the hydrocarbon gases combined no molecules larger than isobutane, (4) by the decrease in chlorinity and salinity characteristic of the freshwater composition of gas hydrates, and (5) by sonic and density logs showing high velocities and corresponding low densities. These are the first demonstrated log responses of cored gas hydrates in oceanic sediment and provide the first measurements of their *in situ* sonic velocity and density.

Gas hydrates were found dispersed in very fine-grained sediments; associated with porous, coarse-grained sediments; occupying fractures; and as a massive unit. Geophysical records off Guatemala do not show strata corresponding to the massive gas hydrate or any other gas-hydrated horizon; only the base of gas hydrate reflecion is laterally continous, and this reflection was observed at many places on the slope, namely at one of the sites occupied (568).

Inorganic geochemical studies on Leg 84 confirm the proposed relationship between low values of salinity and chlorinity and the occurrence of gas hydrates. In all cases where gas hydrates were observed on Leg 84, values of salinity and chlorinity were lower than in sediments where gas hydrates were absent.

Most gas hydrates found were composed of gases from microbial and early diagenetic processes. However, at two sites

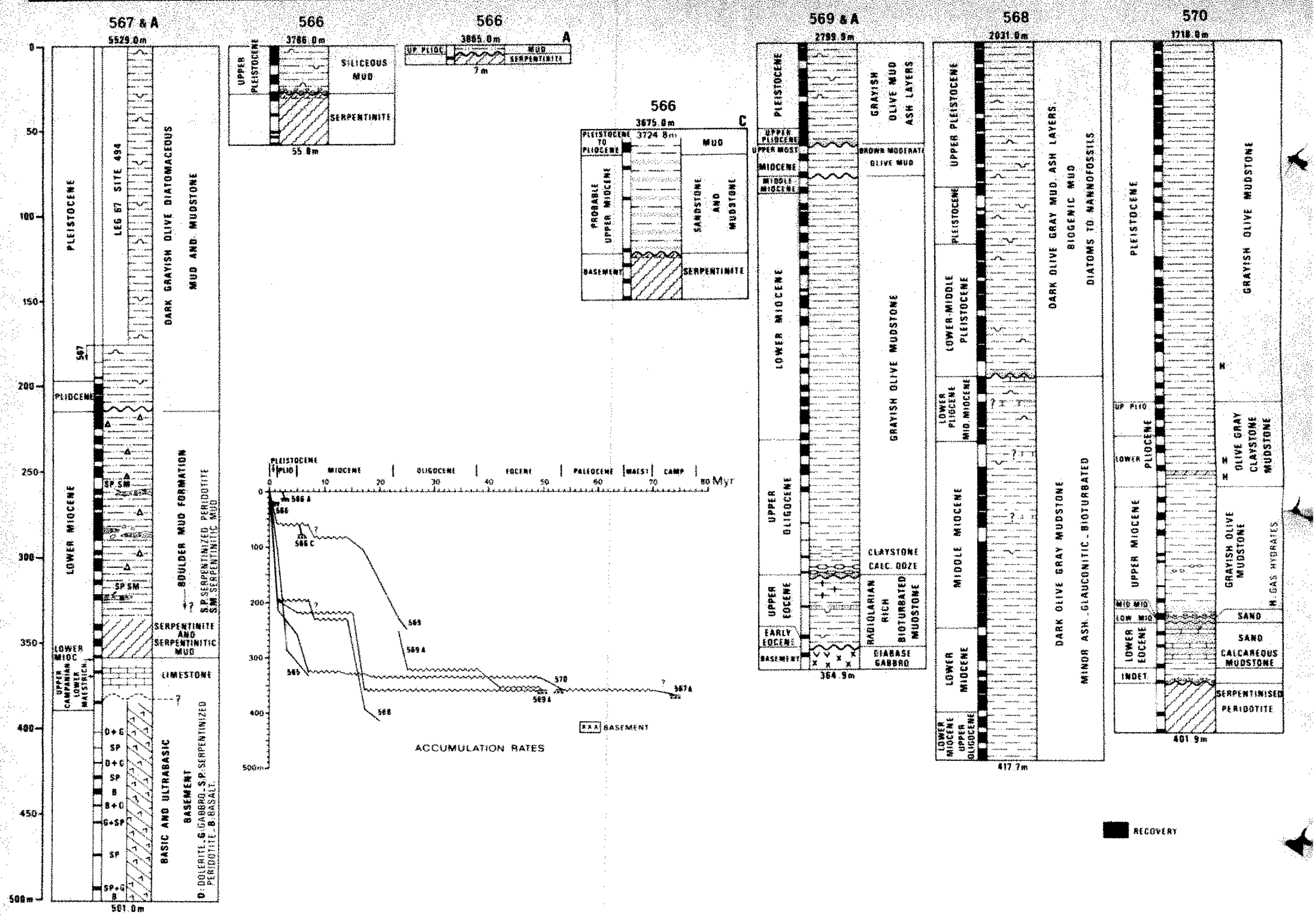


Fig. 3 Simplified stratigraphical columns summarizing the lithostratigraphy, biostratigraphy and recovery of Leg 84 drill sites.

(566 and 570) the gases found associated with serpentinite were probably a product of the thermal breakdown of organic matter because they contained very high concentrations of hydrocarbon gases, particularly of ethane relative to methane, and hydrocarbons at least as large as hexane and heptane.

Conclusions

(1) The DSDP/IPOD Guatemalan transect across the Middle America trench has defined a non-accretionary convergent margin; Neogene subduction has not resulted in the development of an accreted complex; rather, a mass of tectonized ophiolitic rock now faces the subducting Cocos Plate (Fig.3). The ophiolitic rock was emplaced before Eocene time in a previous tectonic system: at the trench the Cocos Plate is of Miocene age.

(2) A horst and graben topography gives the Cocos Plate the appearance of collapsing into the trench, and the adjacent base of the Guatemalan margin is also normally faulted. However, given the 10 cm yr^{-1} plate convergence rate, this extensional topography must be surficial and secondary. The topography of extension in a fundamentally compressional environment can be explained on the seaward slope of the trench by flexure of the Cocos Plate; on the landward slope of the trench it may be explained by a high degree of decoupling across the subduction zone. Decoupling is required to subduct sediment on the lower Cocos Plate beneath the ophiolitic upper plate rather than accreting it. A classical mechanism for such decoupling is overpressured pore fluid such as that measured only a short distance above the Middle America subduction zone. Elevated pore pressure was measured not only in the vicinity of the subduction zone but also in slope sediment across the transect, indicating a general overpressure condition throughout much of the margin.

(3) Beneath the cover of slope sediment, the basement of the Guatemalan margin is an extension of the Central American continent.

(4) The major results of Leg 84 were achieved by drilling below the present slope sediment and into basement. Accretionary complexes may be defined by their geometry in seismic reflection records, but results from most of the convergent margins drilled during the IPOD programme indicate that the accretion observed in seismic records may belong to previous tectonic systems; accretionary periods may alternate with non-accretionary periods. Thus, a seismic record alone may be insufficient to establish the tectonic history or the present tectonic system in a convergent margin. Unless the accretionary complex is sampled, its time of emplacement and relation to the present tectonic system is uncertain.

(5) Gas hydrates are probably dispersed throughout many marine sediment sequences that lie in the zone of gas hydrate stability; they can be detected by seismic techniques when bottom-simulating reflectors appear on seismic records. The occurrence of visible gas hydrates in cores is in part a function of effective porosity and perhaps permeability. Off Guatemala gas hydrate is found finely dispersed in mudstone, as a solid in fractures, in the pore space of sandy ash layers, and in massive form, filling voids of unknown origin. □

1. Moore, J. C. *et al. Nature* **281**, 638 (1979).
2. Aubouin, J. *et al. C.r. hebd. Séanc. Acad. Sci., Paris* **289**, 1215 (1979).
3. Von Huene, R. *et al. Bull. geol. Soc. Am.* **91**, 421 (1980).
4. Aubouin, J., Stephan, J.-F., Renard, V., Roump, J. & Lonsdale, J. *Nature* **294**, 146 (1981).
5. Bourgois, J. *et al. Bull. Soc. géol.* (in the press).
6. Kuijpers, E. *Tectonophysics* **68**, 233 (1980).
7. Schmidt-Effing, R. *Neues Jb. Geol. Paläont. Abh.* **160**, 241 (1980).
8. Azema, J. & Tournon, J. *C.r. hebd. Séanc. Acad. Sci., Paris* **290**, 9 (1980).
9. Seely, D., Cail, P. & Walton, G. in *Geology of Coninental Margins* (eds Burk & Drake) 249 (1974).
10. Kvenvolden, K. & McMenamin, M. *U.S. geol. Surv.* **825**, 1 (1980).

REVIEW ARTICLE

Radar research on thunderstorms and lightning

W. David Rust & Richard J. Doviak

National Severe Storms Laboratory, National Oceanic and Atmospheric Administration, Norman, Oklahoma 73069, USA

Weather radars can resolve many phenomena related to thunderstorms. Echoes from clear air, water and ice condensate, and ionized lightning channels can all be identified. Radial velocity data from two Doppler-radars are synthesized to show a storm mesocyclone with updraft speeds as large as 40 m s^{-1}, moist boundary layer air flowing into the storm, and outflow into the upper troposphere. Maximum wind speeds in tornadoes can be remotely measured using Doppler spectra obtained with a high PRF radar. Doppler radars can detect thunderstorm gusts even when they are in the clear air many tens of kilometres away from the storm, and radar detection of the ionized channels caused by lightning opens new avenues into thunderstorm research.

EVERY year, starting in early spring in the southern states of the USA and ending in mid-summer in the southern provinces of Canada, severe local storms sweep across the Great Plains of North America bringing tornadoes, hail, damaging gusts, and lightning, as well as beneficial rains and cooling outflows. These storms, each a few tens of kilometres in diameter, ingest warm moist air trapped in the first kilometre or two of the atmosphere, convert moisture to rain and sometimes hail, and vent air heated by condensation into the cold upper (9–13 km) troposphere.

Researchers have been observing stormy weather over the past few decades with various instruments. One of the most significant of these is the Doppler weather radar. Radar is the only remote sensing instrument able to expose the internal structure of thunderstorms veiled by clouds that give little evidence of the enclosed hazards. Atmospheric research with radars in the USA is principally supported by: (1) the National Science Foundation, which supports universities directly and through the National Center for Atmospheric Research, (2) the National Oceanic and Atmospheric Administration (NOAA), which maintains and operates a network of weather radars for storm hazard warnings and which contains the Environmental Research Laboratories that support weather radar research having potential operational application, and (3) the Department of Defense, which supports weather radar research through its Air Force Geophysics Laboratory. The French[1] have a small but well organized Doppler weather radar research programme, and the Soviets[2] have long used Doppler radars in weather research. The Canadians[3,4] have a strong research programme to determine precipitation characteristics (such as shape, size, and ice or water) using radars with polarization diversity and dual wavelengths. Recently, the British[5] have shown how dual polarization radar can better identify precipitation type and improve rainfall rate estimates. Research on rainfall and hailfall accumulations, the foundation for many years of weather radar research, continues in the USA[6,7], Switzerland[8], Japan[9], and other countries[10–13]. Recently, aeronomists in the USA[14,15] and Germany[16] have focused large area (for example, 10,000 m^2), long-wavelength (6 m) Doppler radar antennas into the troposphere to study the air motions in cloudless skies. Researchers have also used short-wavelength (10 cm) Doppler weather radars to map winds in the clear air prestorm environment and have obtained results suggesting improvements in prediction of storm location and severity[17,18].

Although lightning is necessary in thunderstorms, only recently have there been coordinated studies to undertake simultaneous measurements of the storm's electrical properties, its precipitation intensity, and its wind fields. One such coordinated effort has been the Thunderstorm Research International Program[19] in which experiments have been conducted in Florida and New Mexico in the USA. Initial results of these and other experiments show correlations between lightning and winds inside the storm[20,21]. Not only does lightning seem to be linked to storm dynamics in the highly convective spring and summer thunderstorms, but it also seems related to the dynamics in weaker winter storms. The recent observations by Brook *et al.*[22] show a correlation between the occurrence of lightning that transfers positive charge to earth and the vertical shear of horizontal wind in the winter thunderstorms along the Hokuriku coast of Japan.

There are many research organizations investigating various aspects of weather phenomena, but at NOAA's National Severe Storms Laboratory (NSSL), located in Oklahoma in the heart of the Great Plains, observations and research are focused on the interactions between storm electricity, precipitation, and wind in severe thunderstorms using: (1) two 10-cm wavelength pulsed Doppler weather radars powerful enough to sense the evolving wind fields in the prestorm environment; (2) the USA's tallest meteorologically instrumented tower; (3) a network of surface-based electrical and meteorological sensors; (4) a unique passive network of radio receivers to map lightning inside clouds and to locate the strike point of cloud-to-gound flashes; and (5) a system to guide instrumented aircraft in their penetration of the intense and turbulent storms.

The continued study of storm electricity, precipitation, and dynamics with recently developed remote sensing instruments could improve our understanding of how thunderstorms organize themselves. Improved predictions and warnings of hazards should result. We now summarize aspects of the use of radar to observe remotely the hazardous phenomena of thunderstorms.

Doppler weather radar

As its beam sweeps through the storm, the weather radar maps the distribution of precipitation by measuring reflectivity factor Z (the product of the number of drops per m^3 and the average sixth power of their diameter expressed in mm (ref. 23)) and generates a three-dimensional 'photograph' of reflectivity fields, typically in 3–5 min. Transmitted radar pulses also reflect from ionized paths created by lightning within the radar beam. Furthermore, radar can detect echoes returned from the small irregularities of the air's temperature and humidity in the optically clear atmosphere[24]. Irregularities of centimetre size and temperature variations as small as 10^{-3} °C produce reflectivity detectable by radar, although invisible to the naked eye. However, the eye can see the scintillation of stars caused by these temperature fluctuations. Echoes from such targets in the first kilometre of the clear atmosphere can be seen in Fig. 1*b*.

Doppler radar adds information through its ability to measure target speeds. Doviak *et al.*[24] have reviewed the principles of

Doppler weather radar, and we highlight those to help interpret the data we present. The motions of targets towards or away from the radar (that is, radial velocities) cause minute shifts (~0.01 p.p.m.) in the microwave frequency of echoes in the same manner as a passing train causes shifts in the observed pitch of its whistle. Because there are usually many millions of targets moving at different velocities, the Doppler radar actually measures a spectrum of radial speeds within its resolution volume, defined by the beamwidth and the pulse length of the transmitted radiation. The resolution volume is an inherent characteristic of the radar for zero antenna scan rate. Because echoes from several tens of transmitted pulses are required to determine the Doppler spectrum, the volume contributing to the spectrum is larger than the resolution volume[25] if the beam scans while receiving echoes. This larger volume is defined as the measurement volume. The width of the radial velocity distribution within the measurement volume can be $<1\ \mathrm{m\ s^{-1}}$ when there is no wind shear or turbulence, or as much as $100\ \mathrm{m\ s^{-1}}$ when a tornado is contained within the volume. The average velocity v and the variance σ_v^2 of the radial velocity distribution are the first and second moments of the Doppler spectrum.

Because the Doppler radar measures only the component of target velocity in the direction of the radar, three or more radars must map Doppler speeds nearly simultaneously from different directions to synthesize the three components of velocity. However, the three components of wind are connected by the mass continuity equation so that only two spaced radars are sufficient to estimate the wind field[26]. Because ice crystals and water drops fall through the air, estimates of their fall speeds relative to the air (terminal velocities) need to be made when these are the tracers of wind. (Note that velocities relative to the ground are fall velocities.) Terminal velocity estimates are usually based on the reflectivity of the hydrometers[27]; larger ones have higher terminal velocities and reflectivities. Refractive index irregularities and lightning channels have zero terminal velocities.

Observations

Storm structure is well illustrated when data from a single radar are displayed on a range height indicator (RHI) that gives vertical cross-sections of Z, v, and σ_v. Figure 1*a*, *b* shows respectively, Z (scaled in decibel units, dBZ) and v, in cross-sections along a 347° azimuth, for a severe thunderstorm. Visible in the Z field (*a*) is an anvil of cloud particles extending to the left (SSE) at ~10 km AGL (above ground level). More of the anvil is evident in the velocity display (*b*) for which the signal-to-noise ratio (SNR) threshold has been reduced to show the motion of the weakly reflective particles, which are probably ice crystals. Storm overhang and strong reflectivity gradients below 2 km AGL on the storm's south side suggest strong inflow into the storm. If we assume that the hydrometers are liquid and have the Marshall–Palmer[28] exponential dropsize distribution often used by radar meteorologists, then the 20 dBZ contour near the ground corresponds to a rainfall rate $<1\ \mathrm{mm\ h^{-1}}$ and 30, 40 and 50 dBZ to ~3, 10 and $50\ \mathrm{mm\ h^{-1}}$, respectively. Often the high reflectivity regions (>50 dBZ) extending to the ground contain wet hail, which causes errors in the estimate of rainfall rate. Melting ice crystals mixed with rain in the region (melting layer) between the freezing temperatures above and warmer temperatures below can also cause a significant error in rainfall rate estimates.

The Doppler velocity field in Fig. 1*b* depicts motions to the right and away (positive velocity) from the radar as well as motion towards (negative velocity) the radar. The storm outflow, assisted by northwesterly winds outside the storm, carried cloud particles in the anvil causing it to grow southeastward towards the radar. In the anvil, air velocities were more negative than $-27\ \mathrm{m\ s^{-1}}$, the Nyquist limit[24] of the radar, so they appear as high positive values (but less than $+27\ \mathrm{m\ s^{-1}}$) and are seen as patches of lightest red embedded in regions of lightest green. The momentum of updraft air carried condensate into the lower stratosphere where negative buoyancy forced air downward and outward around the updraft. This latter motion is evident in Fig. 1*b* above the anvil on the storm's southern side where we see the highest radial velocities.

The velocity data in Fig. 1*b* above the upper edge of the boundary of the reflectivity field shown in Fig. 1*a* are due to antenna sidelobes receiving echoes from the high reflectivity regions when the main beam lobe is pointed above the storm (due to reception of signals from directions away from the antenna axis at angular distances larger than the beamwidth). These falsely mapped velocities do not portray the true wind there. Antenna sidelobes constitute one of the more important limitations to expanding the utility of Doppler radars. However, if there are no strong echoes at a given range in the main or sidelobes, then the Doppler radar can accurately sense the velocities in weak echo regions in the main beam at that range. For example, weakly reflecting targets drifting with the flow in the clear air boundary layer (low altitude, red region in Fig. 1*b* between ranges of 50 and 60 km) trace the air moving into the storm.

Although radial velocity fields from one radar exhibit many interesting storm features, they cannot unambiguously depict the wind. Figure 2*a* shows a horizontal cross-section of a tornadic storm that formed in central Oklahoma[29] on 2 May, 1979. The three components of wind were obtained from radial velocity measurements made by two radars spaced 70 km apart, a reflectivity–terminal velocity relation, and integration of the mass continuity equation using suitable boundary conditions at the ground and the cloud top[30]. The large circulation of air seen in Fig. 2*a* is a thunderstorm cyclone. Because thunderstorm cyclones are much smaller than extra tropical ones or tropical cyclones, including hurricanes, these circulations are called mesocyclones. Not all thunderstorms develop mesocyclones, but when they do so, they have the potential to generate large and long-lived tornadoes as did this one. In all the cases observed, large tornadoes have been preceded by mesocyclones. Mesocyclones have diameters ranging from a few to ~10 km and can extend in height from near the ground to near the tropopause. Tornadoes are smaller scale vortices imbedded within the mesocyclone, and if their centres do not coincide (as in Fig. 2*a*), the tornado circulates around the mesocyclone centre forming a trochoidal path on the ground as the mesocyclone and tornado are translated by the moving storm[31]. Note that the small scale tornado vortex is not resolved in Fig. 2*a*.

The inflow of air from the south-east in the first 2 km AGL, and updrafts as large as $40\ \mathrm{m\ s^{-1}}$ can be seen in Fig. 2*b*. Inflow air lacking precipitation targets is also detected by radar as shown in Fig. 1*b*. Large gradients of reflectivity mark the region of inbound air, and strong outflow occurs in the anvil at 12 km AGL. These are features similar to those seen in the single radar data fields (Fig. 1*a*, *b*). Although the cross-section in Fig. 2*b* shows mostly updrafts, other regions of the storm have downdrafts. The chilled downdrafts are sometimes so

Table 1 Comparison of tornado warning performance, 1977–78

	Present system	Doppler radar alone
Probability of detection*	0.64	0.69
False alarm rate†	0.63	0.25
Critical success index‡	0.30	0.56
Lead time (min)§	2.2	21.4

Approximately 200 cases were examined.
* $X/(X+Z)$; † $Y/(X+Y)$; ‡ $X/(X+Y+Z)$ where X = number of predicted tornadoes that occurred; Y = number of predicted tornadoes that did not occur; Z = number of tornadoes that occurred without a prediction.
§ Lead time between issuing of warning and tornado occurrence.

strong and enduring that they form cooled air masses that extend many kilometres beyond the storm cloud.

Observed winds compared with model predictions

High-speed, high-capacity computers allow us to model numerically thunderstorm evolution. As computers become faster and less expensive and models more computationally efficient, it even becomes feasible to use prestorm environmental parameters (vertical profiles of temperature, moisture and wind) for improved predictions of weather events. Figure 3 shows a comparison of a numerically modelled storm formed in an environment similar to that of an observed storm[32] in which wind at 4 km AGL is from the SSW (199°) at 25 m s^{-1}. Winds are relative to storm motion (14 m s^{-1} at 192° for the model; 18 m s^{-1} at 193° for the observed storm). The heavy contour in Fig. 3*a* corresponds to a reflectivity factor of 38 dB*Z*, assuming the Marshall–Palmer dropsize distribution. Seen in both the modelled and observed storms are the updrafts and downdrafts with reasonable agreement in shape, location and intensity. The updraft blocks the dry environmental air at 4 km and diverts it around the storm whose circulation causes higher speed dry air to intrude into the storm's east side. There the fast-moving dry air deflects and evaporates the condensate, thus forming a notch-like feature in the reflectivity contour, seen in both the model and observations.

Tornado observations

When a tornado is contained within the radar's resolution volume, the Doppler velocity spectrum becomes very broad[33]. When radar targets have small terminal velocities, the Doppler spectra can be analysed to determine the flow of air around the tornado[34,35]. The only scanning of a large tornado (Fig. 4*a*) with a pulsed Doppler radar operating with a Nyquist interval sufficiently large (±91 m s^{-1}) to measure unambiguously the high tornadic wind speeds was in Spring 1981 at NSSL. The Nyquist interval, $\pm\lambda/4T_s$, is determined by radar wavelength λ and the period T_s between transmitted pulses. Targets moving at speeds higher than the Nyquist limits produce signals that are not sampled often enough to resolve unambiguously the Doppler shift. The contour in Fig. 4*a* shows the approximate location and cross-section of 16 consecutive (in range) measurement volumes within which targets produced the 16 Doppler spectra shown in Fig. 4*b*. Each spectrum gives the distribution of velocities within one of the 16 measurement volumes, and each volume has a range extent of ~200 m. These spectra were recorded within a minute of the time of the tornado photograph in Fig. 4*a*. Because the antenna was scanning in azimuth while data were collected to form these spectra, the measurement volume has an azimuthal dimension of 1.5°, although the beamwidth is 0.8° (ref. 25). The 16 volumes are aligned along the

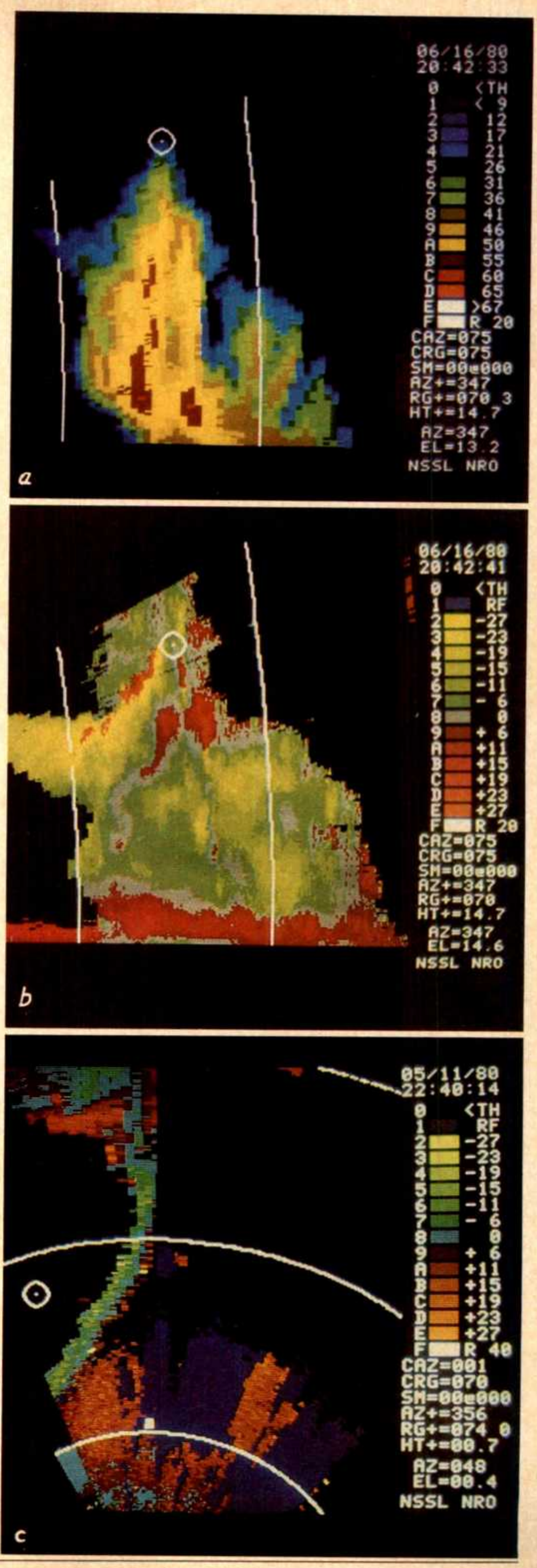


Fig. 1 *a*, Vertical cross-section of reflectivity factor *Z*, a measure of precipitation intensity, through a severe storm that occurred in central Oklahoma on 16 June 1980. The storm top (indicated by dot in white circle) is at 14.7 km AGL. The two shades of blue indicate median values of $10 \log_{10} Z$ (dB*Z*) equal to 17 and 21; the three shades of green correspond to 26, 31 and 36 dB*Z*; the three of yellow to 41, 46 and 50 dB*Z*; and the two of red to 55 and 60 dB*Z*. The relation between dB*Z* values and rainfall rate is given in the text. The radar location is to the left, and the vertical arcs are at 60 and 80 km from the radar. *b*, Radial velocity field in the same vertical section as *a*. Red (green) coloured areas are radial velocities away from (towards) the radar. The shades from dark to light correspond to median radial velocities of 6, 11, 15, 19, 23 and 27 m s^{-1}. The grey hue is for a median value of 0 m s^{-1}. Differences in storm outline from that in *a* are discussed in the text. *c*, Radar depiction of radial velocities in a bow gust line produced by thunderstorm outflow at ground level. Colour code is the same as in *b*. This is an azimuthal sweep through the gust at a constant elevation angle of 0.4°, and the arcs are at 40 and 80 km range.

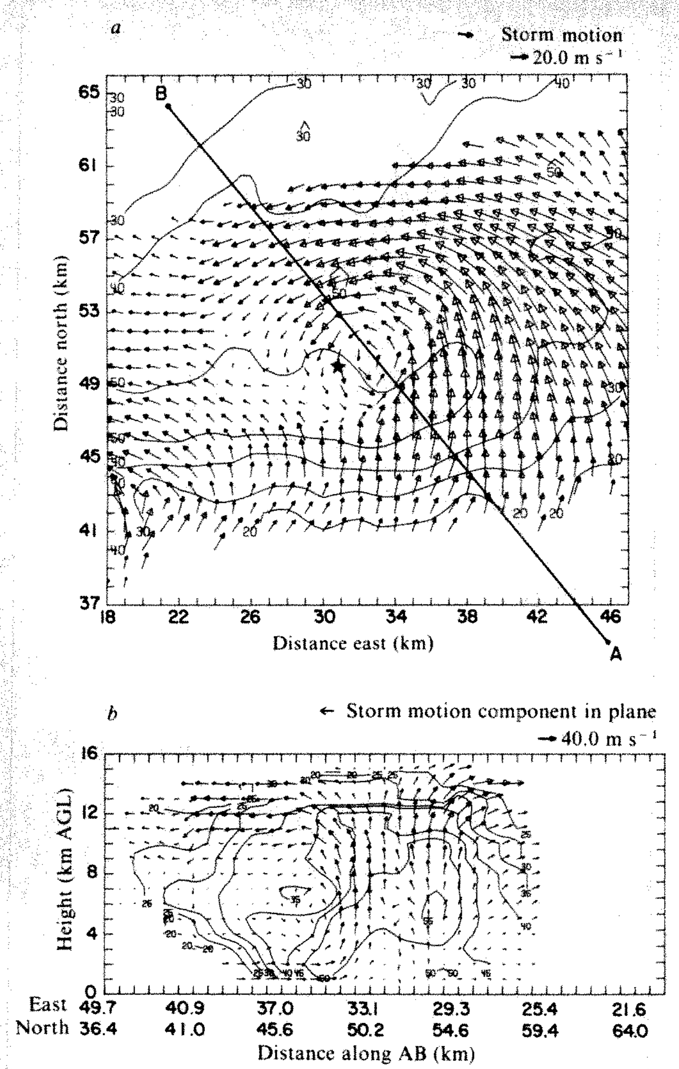


Fig. 2 Contours of reflectivity factor in dBZ and the wind field in a horizontal cross-section at 2 km AGL (a) and a vertical cross-section (b) of a tornadic thunderstorm. Winds are in the coordinate system moving with the storm. Storm motion speed and direction are shown. Line AB in a is the location of the vertical cross-section. Distances are from a radar at Roman Nose State Park, Oklahoma[29]. ★, The tornado location in a. The arrow of indicated velocity at top right of each plot gives speed which is proportional to arrow length (E. Brandes and B. Johnson, personal communication).

beam, but only two or three of these cross the tornado's circle of maximum wind. In Fig. 4b the widest spectra denote the location of the tornado. Maximum tornado wind speeds can be estimated from these data. At this time the radar-measured maximum winds were ~90 m s^{-1}. Acquiring *in situ* data in and around the tornado is very difficult, but Doppler radar provides a remote and relatively safe method of surveying storms for tornadoes.

Although in principle two or more Doppler radars are required to depict circulations inside storms, a single Doppler radar can indicate circulations, such as mesocyclones, that have nearly circular symmetry. The circulation signature of large mesocyclones shows itself remarkably well on colour displays (see cover of ref. 36). Because mesocyclones are often several kilometres in diameter (see Fig. 2a), their circulations can be resolved to ranges of ~200 km with radars having beam widths <1°.

A test was conducted in Oklahoma to determine how tornado warnings are improved over the present non-Doppler warning system when Doppler information alone is used to forecast tornadoes. Table 1 compares several measures of effectiveness of tornado warning using Doppler radar alone with present techniques which rely heavily on visual sightings. The critical success index (CSI)[37] gives credit for a high probability of detection and low false alarm rate, with a CSI of 1.0 being perfect.

Doppler radar can increase the average lead time before tornado occurrence by 20 min, thus offering the population a longer time to take cover if warnings are communicated properly. Although the probability of detection at the instant of issuance of warning is not significantly improved, the false alarm rate is significantly decreased with Doppler radar. Therefore, Doppler radar can reduce the overwarning inherent in the present system even while giving longer lead times, and warning areas may be reduced substantially by precise location of the storms containing mesocyclones.

Gust fronts

A gust front is the leading edge of cool air produced when a raincooled downdraft turns into an outward moving, ground-based flow. The front is marked by shifts (shear) in the wind, both in the vertical and horizontal directions. A gust front can propagate in the clear air many tens of kilometres away from the thunderstorm that caused it and yet harbour shear forces that can be destructive to aircraft, especially when a flight crew is unaware of its presence. The wind behind the front is usually turbulent and attached to the storm's downdraft[38]. But in some cases, as shown in Fig. 1c, the outflow air can detach itself from the storm and propagate away from it. Gust fronts often appear as thin lines on radar displays Fig. 1c, and reflectivity is so weak (<10 dBZ) that some radars fail to detect it. However, moderately sensitive Doppler weather radars can sense reflectivities as low as −10 dBZ at ranges <60 km.

Figure 1c shows the Doppler velocity field of a gust produced by a thunderstorm, which can be partially seen at the top of the figure. The radar beam elevation was fixed at 0.4° while the radar scanned in azimuth from 330° to 45°. Note the abrupt change in radial velocity across the gust line. The wind on the radar side (south-east) of the line shows a component away from the radar, and just beyond the line, towards the radar. The region of approaching velocities along the gust is rather narrow. The air flow immediately behind the 4-km wide band of approaching air is again southerly, although it is not evident in Fig. 1c because the SNR threshold eliminated the weaker echoes behind the gust. The dark blue area near the 40 km range arc indicates regions where the velocity data are contaminated with echoes from distant (>150 km) storms along the same bearing. A time sequence of radar images of the thunderstorm and gust suggests that this outflow propagated more than 80 km with a relatively constant velocity of 13 m s^{-1}. Fortunately, the gust passed the 444-m tall KTVY television tower, instrumented by NSSL, so that *in situ* measurements of the wind and temperature were obtained.

The vertical profiles of the wind and temperature in a south-east to north-west cross-section obtained from a composite of radar and tower data are shown in Fig. 5. The data show this solitary gust to be a cylindrical pool of cool air that propagates southeastward. This highly localized pool of steadily moving cool air appears to be detached from the storm and propagates in a stable boundary layer with little mixing of its air with the environment so that the air behind the line quickly returns to ambient conditions. This solitary gust seems to have a remarkably long lifetime (>1 h) as judged from time lapse photography.

Turbulence

Evident in Figs 2 and 5 (and in the visual images of cumulus clouds) are small scale eddies, perturbations in the larger scale circulations of storm winds. Radars with finer resolution than the nearly 2 km resolution of Fig. 2 and *in situ* point measurements made with aircraft often show small scale thunderstorm

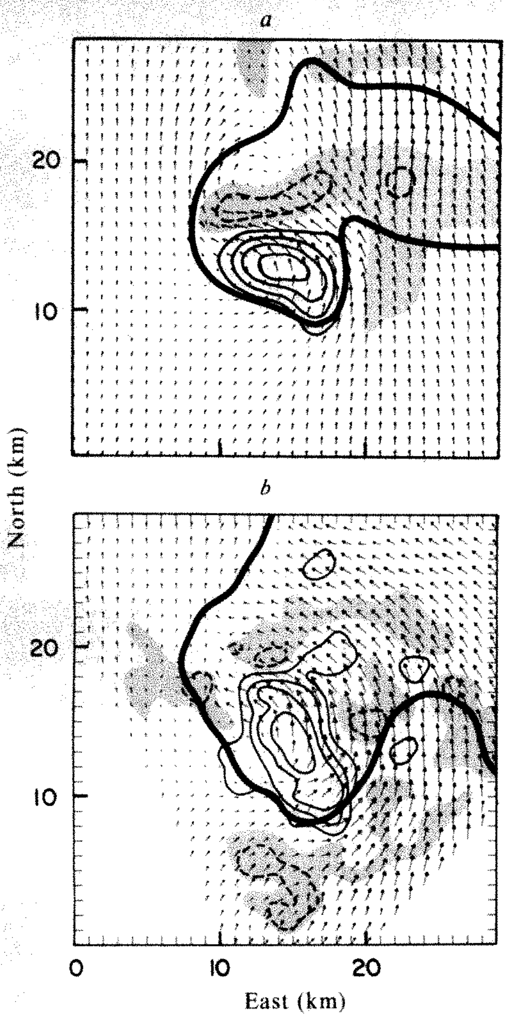


Fig. 3 Horizontal cross-section of modelled storm winds (*a*) and winds synthesized from observed Doppler velocities (*b*) in a plane at 4 km AGL (ref. 32). Updraft velocities (solid lines) and downdraft velocities (dashed lines) are contoured in 5 m s^{-1} increments. Shaded areas are downdraft regions stronger than -1 m s^{-1}. The heavy solid contour in *a* encloses a region of liquid water concentration >0.5 g kg^{-1}. In *b* the heavy contour is a 30 dB*Z* reflectivity factor. Wind vectors are scaled such that one grid interval is 20 m s^{-1}.

wind to be chaotic. If these turbulent eddies are intense, they can be destructive to aircraft.

The width, σ_v, of the Doppler velocity spectrum is normally a measure of turbulence within the measurement volume, although shear, antenna scan rate, and differential fall speeds contribute to a lesser degree to the measured value[24,39]. Although an aircraft can record turbulence in narrow volumes along its flight path, and a radar resolves it within its usually much larger measurement volume, Bohne[40] has shown good correlation between levels of turbulence estimated from aircraft and radar. Analysis of a larger amount of data by Lee[41] suggests a strong correlation between σ_v and *in situ* aircraft measurements of turbulence. His data show that when aircraft-measured gust velocities exceed 6 m s^{-1} (moderate or severe turbulence), σ_v exceeds 5 m s^{-1} within 1 km of the radar measurement volume. However, not all storm regions of large σ_v produced turbulence detected by aircraft. Furthermore, in over 50 thunderstorm flights when σ_v was <4 m s^{-1}, the aircraft experienced only light turbulence.

Only meagre results relate radar-estimated turbulence to storm wind fields; nevertheless, present evidence suggests that the most intense turbulence in a tornadic thunderstorm occurs along the edges of high speed (≥ 20 m s^{-1}) updrafts[42]. The most intense turbulence is usually found in the midlevels of the storm where there is large shear between updrafts and downdrafts[43].

Rainfall measurements

Although radar estimates of rainfall rate R obtained from Z are typically in error by a factor of up to two, radar has the advantage of surveying remotely vast areas and making millions of measurements in minutes. A sufficiently dense network of rain gauges can measure the rainfall more accurately and reliably. However, no matter how accurate the gauges are for point measurements (typically 5–10%), their usefulness for determining the rainfall integrated over a large area is limited by gauge density and the spatial variability in the rainfall[44]. The cost of a rain gauge network to match the radar's capabilities

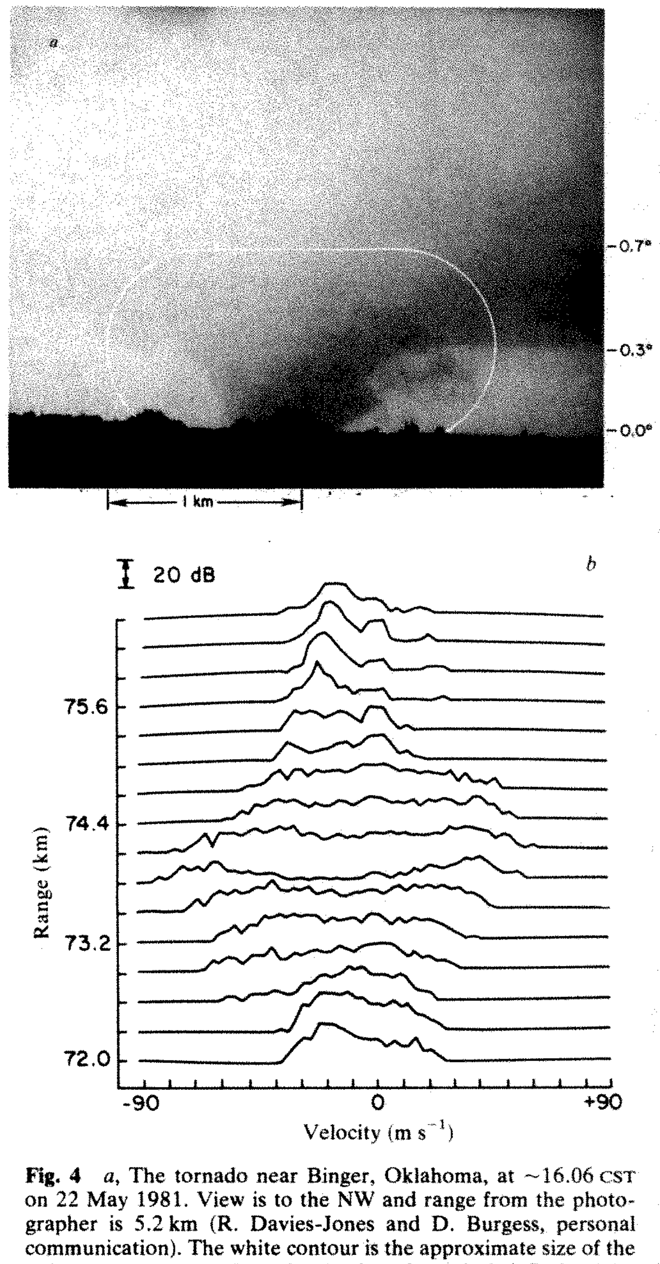


Fig. 4 *a*, The tornado near Binger, Oklahoma, at ~16.06 CST on 22 May 1981. View is to the NW and range from the photographer is 5.2 km (R. Davies-Jones and D. Burgess, personal communication). The white contour is the approximate size of the radar measurement volume for the data shown in *b*. *b*, Reflectivity spectral densities (in dB units) versus Doppler (radial) velocity at 16 range locations in the tornado's circulation. Beam elevation angle is 0.3°. The nearly horizontal lines at each end of the spectra are thresholds at 10 dB above the receiver noise level and correspond to a reflectivity spectral density of ~3.4 mm^6 m^{-3} m^{-1} s. The vertical scale is 20 dB per division (L. Hennington and D. Zrnic', personal communication).

in spatial continuity and rates of data sent to a central location would be prohibitive. Therefore, radar and widely-spaced rain gauge data are combined to produce rainfall maps. Inherent in this procedure are the assumptions that, despite the possible error in absolute radar measurement of R, the radar accurately measures the spatial variability of R and a valid calibration can be made between gauge and radar measurements at each gauge location. With these assumptions such rainfall maps approach the point accuracy of a dense gauge network[45,46].

The distribution of dropsize, $N(D)$, is of central importance in determining R. Operational weather radars measure the single parameter Z, and meteorologists often estimate R by using an assumed exponential dropsize distribution (see ref. 28) in which only one parameter is estimated from Z. In general, however, specification of $N(D)$ requires knowledge of several independent parameters to describe it, so that accurate estimates of R and Z can be made. It has been suggested that when $N(D)$ is described by a two-parameter distribution, pairs of independent measurements of two remotely sensed properties, such as attenuation and reflectivity, can be used for a better estimation of R (ref. 47).

A dual polarization radar[48] can use echo intensity information contained in two orthogonally polarized waves to estimate both parameters of a two-parameter $N(D)$. The basis for the dual polarization scheme is the observation that drops, falling at their terminal velocities, are not spherical but have an oblate spheroidal shape, so that we expect larger echo power for horizontally polarized waves. It can be shown that the ratio Z_H/Z_V of reflectivities obtained with horizontally and vertically polarized waves is an accurate measure of the spatial variability of R, even if the absolute radar calibration is in error, if the assumption of exponential dropsize is valid. Then rain gauges, even though sparsely spaced, can be used to scale the radar estimates of R. Although results given by Seliga *et al.*[49] suggest that a dual polarization radar can give better estimates of rainfall rate, more research is needed to evaluate the technique for operational applications.

Because the ratio Z_H/Z_V is related to the shape of the scatterers, polarization analysis can also be used to identify precipitation type (rain, hail, graupel, and so on) when this ratio and absolute measurements of Z are categorized into levels expected for various types of hydrometers[5].

Lightning studies using radar

Lightning flashes occur as either cloud-to-ground (CG) or intracloud (IC). A CG flash is comprised of one or more return strokes, which have large electrical currents in the channel between the cloud and ground, while all flashes not reaching the ground are termed IC. (See ref. 50 for a general description of lightning and its measurement.) Both IC and CG flashes create ionized channels that carry electrical currents. Backscattering from these ionized channels is sufficient to produce radar echoes.

Radar detection of lightning has been known for many years[51–58]. Although some early studies included estimates of values for lightning parameters such as electron density, it was not until 1972 that a theory of lightning echoes from return stroke channels was proposed[59]. Since then, new measurement techniques and theories have been developed to help unravel the complexities of lightning echoes.

Characteristics of lightning echoes

In a few early studies of lightning echoes, attempts were made to relate the echoes to known lightning processes by comparing the sferics from lightning (the transient electromagnetic field radiated by a flash) with the radar echoes of the same flashes[57,60]. More recently, coincident measurements of lightning echoes and electrostatic field changes, which allow reasonably conclusive identification of flash type and inter-flash events, showed that lightning echoes from a CG flash often have sudden increases in amplitude associated with return strokes followed by a decay to pre-flash levels[61] (Fig. 6). This feature has also been noted during some interstroke intervals. The radar echoes from an IC flash usually have a relatively broad peak with variations superimposed, which may be seen in Fig. 6*b* during the first part (0.6–1.15 s) of the flash. The electric field changes, (R in Fig. 6) are typical of return strokes. This flash was independently verified by a lightning strike location system as a CG with at least eight return strokes. Note that abrupt increases in the lightning echoes in a particular range interval ΔR may not always coincide with the return stroke field change.

Comparison of lightning echoes with electric field changes suggests that a continuing discharge process often keeps the channels ionized[62,63]. We define the duration of lightning echoes as the time interval from the instant the echoes are first discernible until they decay back to the pre-discharge amplitude. This duration forms a lightning echo event. Figure 6 shows at least nine echo events. Typical lightning echo events have durations of ~200–300 ms with extreme values of 10 ms and 3 s having been observed.

The rise time (the time for the radar received signal to reach peak amplitude) of a lightning echo event is generally a few tens of milliseconds. Although it was first thought that the rise time indicates increasing current flow in the channel[57], more recent analysis[64] shows that observed rise times agree well with ones determined theoretically, when the lightning is modelled as an ionized channel or streamer propagating across the antenna beam at a rate of $\sim 100\ \mathrm{km\ s^{-1}}$. Propagation speeds of this order of magnitude for streamers within clouds have been inferred by several investigators, using different measurement techniques[65,66]. When streamers propagate radially along the beam, their echoes can be seen developing in contiguous resolution volumes. Radar observations of several thousand lightning flashes in Oklahoma storms have yielded a maximum propagation speed of $\sim 250\ \mathrm{km\ s^{-1}}$, and streamer propagation speeds of $>100\ \mathrm{km\ s^{-1}}$ have been measured within small storms[62].

The radar cross-section, RCS, of lightning has been estimated both theoretically and experimentally. The lightning RCS depends on two factors: (1) the size and number of reflecting

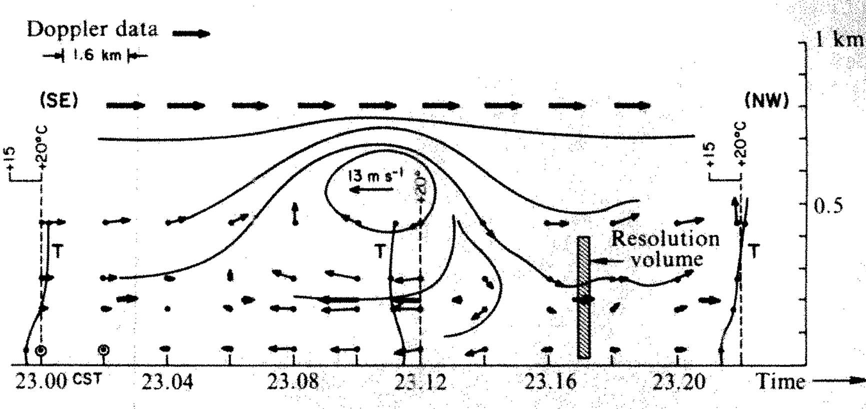


Fig. 5 SE–NW cross-section of a gust. Lightly traced arrows are the tower measured winds projected onto this plane, bold arrows the Doppler velocity. T, temperature profiles. The hatched area shows the radar's resolution volume cross-section in the plane of the beam axis.

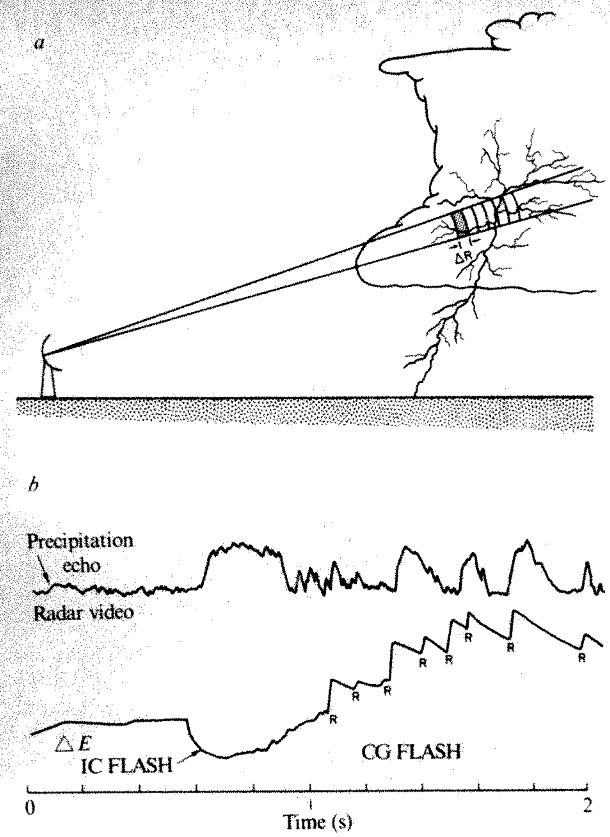


Fig. 6 Depiction of a technique (*a*) for observing lightning and an example (*b*) of lightning echoes for a long flash that began as an IC and then became a CG. The radar antenna typically is held at a fixed position to acquire lightning echoes. The stippled region in *a*, ΔR, represents one of the radar's resolution volumes in which the time variation of echo intensity may be studied. Top trace of *b* is time history of lightning echo intensity in a resolution volume at 86 km and 4 km altitude. Although depicted as a continuous line, the trace is actually a series of points separated in time by the period T_s ($\simeq$1 ms) between transmitted pulses. Precipitation echo denotes pre-flash radar echo intensities due to precipitation. The bottom trace is the electrostatic field change at the radar site due to the same flash. (Reprinted from ref. 21.)

lightning channel elements within the radar resolution volume and their orientations relative to the incident wave, and (2) the channel plasma characteristics. The initial theoretical work was principally done by Dawson[59], who assumed that only those ionized channels perpendicular to the beam axis contributed significantly to lightning echoes, and Divinsky[67], who considered the return echoes irrespective of channel orientation. Both treated the limiting cases of low density and overdense plasma (the plasma is considered overdense when the radar signals are reflected from it as if it were a metallic conductor). More recently Mazur[64] estimated the RCS as a function of channel orientation relative to the incident wave and any level of electron density and temperature by treating lightning elements as dielectric cylinders of finite length. Experimentally determined estimates of the lightning RCS reported by Mazur[64], Holmes *et al.*[62], and Zrnic' *et al.*[63] are reasonably consistent. Holmes *et al.*[62] conclude that the in-cloud portion of their flashes, which were at altitudes above 6 km, had channels with diameters <2 cm and electron densities of $\sim 10^{16}$ m^{-3}.

Lightning and storm structure

The fact that radars of wavelength $\geqslant$10 cm can easily locate lightning, even in intense precipitation, out to ~300 km makes them useful in the study of lightning relative to storm structure. The use of Doppler radar to observe lightning is a promising new area of research.

A phenomenon of long-standing interest and controversy is the 'rain gush', which is a dramatic increase in rainfall soon after lightning. The controversy has arisen as a result of different interpretations of the gush: one is that lightning causes the rapid growth of precipitation[68], another is that the precipitation, in electrostatic levitation before the flash[69,70], is merely released by the change in the electric field caused by the flash. Other interpretations have also been suggested.

Recently two radars have been used by Szymanski *et al.*[71] to search for a relationship between rain gushes and lightning. A 3-cm wavelength radar was used to observe precipitation reflectivity while an 11-cm wavelength radar was used to record the location of lightning echoes. They report that of more than 150 flashes observed by radar, only one flash, which occurred in very low precipitation reflectivities, seemed to cause precipitation growth in the volume around the lightning. Further such observations with radar are needed to provide more definitive information concerning the fundamental question of electrical effects on precipitation.

Zrnic' *et al.*[63] found, as had Szymanski *et al.*[71], that lightning in heavy precipitation usually causes no discernible change in precipitation reflectivity. However, Zrnic' *et al.* did find that lightning altered precipitation reflectivity of low Z and suggested that these were due to changes in orientation brought about by electrostatic field changes in these regions. These changes occurred within 0.1–0.5 s after the flash began. McCormick and Hendry[3] give convincing evidence that both aerodynamic forces and the storm's electric field control the orientation of ice crystals.

We have combined measurements of precipitation, using our 10-cm wavelength Doppler radar, and lightning echoes, using our 23-cm wavelength radar, to study the coevolving lightning and precipitation structure of large storms. In a squall line (that is, storm cells along a line) on 21 June 1980, over 1,000 lightning flashes were observed during a 44-min period. The storm cells were aligned nearly radially from the co-located radars, and lightning flashes from several cells were recorded. For this one squall line, the following were noted[64]: although lightning is observed throughout the storm, most of the lightning occurs close to the leading edge (relative to storm movement) of the heavy precipitation; the incidence of long lightning flashes that propagate between growing and dissipating cells increases as the precipitation echo intensity decreases in the dissipating cell.

The length of lightning flashes determined from the radial extent of lightning echoes varies greatly. Holmes *et al.*[62] report lengths up to 2 km in small, isolated mountain thunderstorms. Ligda's[52] photographs show lightning with lengths up to ~160 km in large storms. Most lightning that we observe in Oklahoma is several tens of kilometres long, and we have

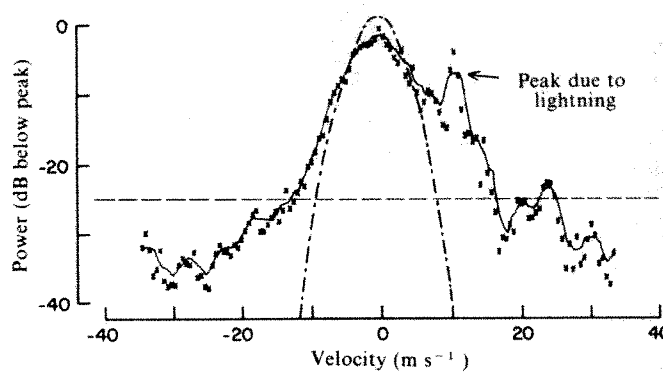


Fig. 7 Reflectivity spectral density for vertical fall velocities from precipitation and lightning at 7.1 km. The dash-dotted curve is the estimated true fall speed spectrum. The solid curve is a five-point running average of the data (×), which were obtained by averaging 12 spectra. Powers below the dashed line and the peak at 24 m s^{-1} are considered to be unreliable, so the intersection of this line and the observed spectrum yields the assumed, reliable maximum fall velocity of ~14 m s^{-1}.

observed lengths in excess of 100 km. These very long flashes have occurred in severe storms or along a line of storms.

Lightning often appears visually to occur in closely-spaced sequences, suggesting that the first flash somehow causes others. Mazur[64] has studied this phenomenon with radar. He observed lightning events characterized by a time interval of <200 ms between them and with a separation in the range of only a few kilometres. He has recorded these in squall lines both in the East Georgia region of the USSR and in the Great Plains of the USA. Mazur termed these 'associated discharges' because of the apparent association between the seemingly separate lightning flashes. It is, of course, possible that different branches of the same flash propagated by chance into the beam to give the observed apparent association. However, Mazur's analysis of lightning in an intense storm showed that this possibility was <5%. From this, he concluded that there may be processes that result in one flash being initiated by another.

Zrnic' *et al.*[63] have used a vertically pointing, 10-cm wavelength Doppler radar to study interactions between lightning and precipitation. Figure 7 shows an example of the reflectivity spectral density of lightning and precipitation observed simultaneously by Doppler radar. Reflectivity spectra obtained with vertically pointed radars have been used to determine the drop size distributions[72,73]. Unless the vertical air velocity and turbulence are estimated with precision better than $1\ \mathrm{m\,s^{-1}}$, large errors in drop size distribution will result. Although several methods have been proposed, none will provide correct values of vertical air motion[74]. If lightning produces ionized channels that permeate the measurement volume, these channels should be perfect tracers of the updraft and turbulence because they have zero terminal velocity.

A subjectively estimated true reflectivity spectrum for the hydrometers is also plotted on Fig. 7 to show how the spectrum depends on velocity. The true spectrum has an upper speed limit equal to the mean fall speed of the smallest detectable drops. Departures of the observed spectrum from the true is caused by turbulence, window broadening (a signal processing effect), possibly returns received via sidelobes, or other non-meteorological effects. Assuming that the ionized channels drift with the wind, Zrnic *et al.*[63] estimated the updraft velocity to be $10\ \mathrm{m\,s^{-1}}$, and so the maximum reliable terminal velocity estimate for the largest hydrometer is $\sim 24\ \mathrm{m\,s^{-1}}$. Such a high velocity was attributed to hail, which indeed fell in the vicinity of the radar. Although this newly tested Doppler radar technique shows potential to determine true updraft velocity and to estimate the radar volume-averaged drop size distribution, much more research is required to ascertain the accuracy and utility of this method.

Concluding remarks

The usefulness of weather radar as a remote probe of large and often severe thunderstorms continues to expand. When used in combination with other sensors, it offers us the opportunity to learn more about the complex interrelations between wind, water, and electricity in storms. Much of this understanding can then, in turn, be applied to the many weather-related problems that affect numerous aspects of our daily lives.

We thank our scientific colleagues, many of whom we could not cite, for their contributions to our understanding of thunderstorms. We thank the staff of NSSL, whose diligent work enabled us to accomplish our research. We particularly acknowledge many stimulating discussions with Vladislav Mazur. Our participation in this work was supported in part by the Atmospheric Science Program of the United States Office of Naval Research, the Severe Storms and Local Weather Research Office of NASA, and the Federal Aviation Administration of the US.

1. Nutten, B. *et al. Trans. Geosci. Electr., Spec. Iss.* **GS-17**, 281–288 (1979).
2. Chernikov, A. A., Mel'nichuk, Y. V., Pinus, N. Z., Shmeter, S. M. & Vinnichenko, N. K. *Radio Sci.* **4**, 1257–1259 (1969).
3. McCormick, G. C. & Hendry, A. *IEEE Trans. Geosci. Electr., Spec. Iss.* **GE-17**, 142–150 (1979).
4. Humphries, R. G. & Barge, B. L. *IEEE Trans. Geosci. Electr., Spec. Iss.* **GE-17**, 190–195 (1979).
5. Hall, M. P. M., Cherry, S. M., Goddard, J. W. F. & Kennedy, G. R. *Nature* **285**, 195–198 (1980).
6. Green, D. R., Hudlow, M. & Johnson, E. R. *Preprints, 19th Conf. on Radar Meteorology* 470–479 (American Meteorological Society, Boston, 1980).
7. Sengupta, S., Smith, P. L., Dennis, A. S. & Doneaud, A. A. *Preprints 19th Conf. on Radar Meteorology* 461 (American Meteorological Society, Boston, 1980).
8. Waldvogel, A., Federer, B. & Hogl, D. *Preprints, 19th Conf. on Radar Meteorology* 493–498 (American Meteorological Society, Boston, 1980).
9. Tanaka, H., Shinozuka, T., Nakamura, K., Koike, K. & Kuroiwa, H. *IEEE Trans. Aerospace Electr. Syst.* **AES-16**, 567–580 (1980).
10. Zawadski, I. *Preprints, 20th Conf. on Radar Meteorology* 402–407 (American Meteorological Society Boston, 1981).
11. Aronsson, M. *Preprints, 19th Conf. on Radar Meteorology* 74–75 (American Meteorological Society, Boston, 1980).
12. Puhakka, T. M. *Preprints, 19th Conf. on Radar Meteorology* 74–75 (American Meteorological Society, Boston, 1980).
13. Calheiros, R. V. & Antonio, M. *Preprints, 19th Conf. on Radar Meteorology*, 727–729 (American Meteorological Society, Boston, 1980).
14. Green, J. L., Gage, K. S. & Van Zandt, T. E. *IEEE Trans. Geosci. Electr., Spec. Iss.* **GE-17**, 262–280 (1979).
15. Balsley, B. B., Ecklund, W. L., Carter, D. A., Johnston, P. E. & Riddle, A. C. *Preprints, 20th Conf. on Radar Meteorology* 1–6 (American Meteorological Society, Boston, 1981).
16. Röttger, J. & Schmidt, G. *IEEE Trans. Geosci. Electr., Spec. Iss.* **GE-17**, 182–189 (1979).
17. Doviak, R. J., Rabin, R. & Koscielny, A. *Preprints, 20th Conf. on Radar Meteorology*, 546–553 (American Meteorological Society, Boston, 1981).
18. Koscielny, A., Doviak, R. J. & Rabin, R. *J. appl. Met.* **21**, 197–210 (1982).
19. Pierce, E. T. *Bull. Am. met. Soc.* **57**, 1214 (1976).
20. Lhermitte, R. & Krehbiel, P. R. *IEEE Trans. Geosci. Electr.*, **GE-17**, 162–171 (1979).
21. Rust, W. D., Taylor, W. L., MacGorman, D. R. & Arnold, R. T. *Bull. Am. met. Soc.* **62**, 1286–1293 (1981).
22. Brook, M., Nakano, M. Krehbiel, P. & Takeuti, T. *J. geophys. Res.* (in the press).
23. Battan, L. J. *Radar Observation of the Atmosphere* (University of Chicago Press, 1973).
24. Doviak, R. J., Zrnic', D. S. & Sirmans, D. S. *Proc. IEEE* **11**, 1522–1553 (1979).
25. Zrnic', D. S. & Doviak, R. J. *IEEE Trans. Aerospace Electr. Syst.* **AES-12**, 551–555 (1976).
26. Ray, P. S. *et al. J. appl. Met.* **14**, 1521–1530 (1975).
27. Atlas, D., Srivastava, R. C. & Sekhon, R. S. *Rev. Geophys. Space Phys.* **2**, 1–35 (1973).
28. Marshall, J. S. & Palmer, W. *J. Met.* **5**, 165–166 (1948).
29. Alberty, R. L., Burgess, D. W., Hane, C. E. & Weaver, J. F. *SESAME 1979 Operations Summary* (U.S. Department of Commerce, NOAA/ERL, Boulder, 1979).
30. Ray, P. S., Ziegler, C. L., Bumgarner, W. & Serafin, R. J. *Mon. Weath. Rev.* **108**, 1607–1625 (1980).
31. Ray, P. S., Hane, C. E., Davies-Jones, R. P. & Alberty, R. L. *Geophys. Res. Lett.* **3**, 721–723 (1976).
32. Klemp, J., Wilhelmson, R. & Ray, P. S. *J. Atmos. Sci.* **38**, 1558–1580 (1981).
33. Zrnic', D. S. & Doviak, R. J. *J. appl. Met.* **14**, 1531–1539 (1975).
34. Zrnic', D. S., Doviak, R. J. & Burgess, D. *Q. Jl R. Met. Soc.* 707–720 (1977).
35. Zrnic', D. S., Istok, M. *J. appl. Met.* **19**, 1405–1415 (1980).
36. Doviak, R. J. (ed.) *IEEE Trans. Geosci. Electr., Spec. Iss.* **GE-17**, (1979).
37. Donaldson, R. J. Jr, Dyer, R. M. & Kraus, M. J. *Preprints, 9th Conf. on Severe Local Storms*, 321–326 (American Meteorological Society, Boston, 1975).
38. Goff, R. C., Lee, J. T. & Brandes, E. A. *Rep. No. FAA-RD-77-119*, (U.S. Department of Transportation, Washington DC, 1977).
39. Istok, M. *Preprints, 20th Conf. on Radar Meteorology*, 454–458 (American Meteorological Society, Boston, 1981).
40. Bohne, A. R. *Preprints, 20th Conf. on Radar Meteorology*, 446–453 (American Meteorological Society, Boston, 1981).
41. Lee, J. T. *Final Rep. No. FAA-RD-77-145* (FAA, Washington, DC 20590, 1977).
42. Zrnic', D. S. & Lee, J. T. *J. Aircraft* (in the press).
43. Frisch, A. S. & Strauch, R. G. *J. appl. Met.* **15**, 1012–1017 (1976).
44. Huff, F. A. *J. appl. Met.* **18**, 1316–1326 (1970).
45. Brandes, E. *J. appl. Met.* **14**, 1339–1345 (1975).
46. Wilson, E. & Brandes, E. *Bull. Am. met. Soc.* **60**, 1048–1058 (1979).
47. Ulbrich, C. W. & Atlas, D. *J. geophys. Res.* **83**, 1319–1325 (1978).
48. Seliga, T. A. & Bringi, V. N. *Radio Sci.* **13**, 271–275 (1978).
49. Seliga, T. A., Bringi, V. N. & Al-Khatib, H. H. *Preprints, 19th Conf. on Radar Meteorology*, 440–445 (American Meteorological Society, Boston, 1980).
50. Uman, M. A. *Lightning* (McGraw-Hill, New York, 1969).
51. Ligda, M. H. *Bull. Am. met. Soc.* **31**, 279–283 (1950).
52. Ligda, M. H. *J. atmos. terr. Phys.* **9**, 329–346 (1956).
53. Browne, I. C. *Nature* **167**, 438 (1951).
54. Miles, V. G. *Nature* **170**, 365–366 (1952).
55. Miles, V. G. *J. atmos. terr. Phys.* **3**, 258–262 (1953).
56. Hewitt, F. J. *Proc. phys. Soc. Lond.* B **66**, 895 (1953).
57. Hewitt, F. J. *Proc. phys. Soc. Lond.* B **70**, 961 (1957).
58. Atlas, D. *Recent Advances in Atmospheric Electricity* (ed. Smith, L. G.) 441–459 (Pergamon, New York, 1959).
59. Dawson, G. A. *J. geophys. Res.* **7**, 4518–4528 (1972).
60. Galperin, S. M. & Stepanenko, V. D. *Atmos. Elect.* 170–177 (1976).
61. Szymanski, E. W. & Rust, W. D. *Geophys. Res. Lett.* **6**, 527–530 (1979).
62. Holmes, C. R., Szymanski, E. W., Szymanski, S. J. & Moore, C. B. *J. geophys. Res.* **85**, 7517–7532 (1980).
63. Zrnic', D. S., Rust, W. D. & Taylor, W. L. *J. geophys. Res.* (in the press).
64. Mazur, V. thesis, Univ. Oklahoma, (1981).
65. Brook, M. & Ogawa, T. *Lightning* Vol. 2 (ed. Golde, R. H.) 191–230 (Academic, London, 1977).
66. Proctor, D. E. *J. geophys. Res.* **86**, 4041–4071 (1981).
67. Divinsky, L. J. *Atmos. Elect.* 177–185 (1976).
68. Vonnegut, B. & Moore, C. B. *Physics of Precipitation* in Monograph No. 5 (American Geophysical Union, 1960).
69. Schonland, B. F. J. *The Flight of Thunderbolts* (Clarendon, Oxford, 1950).
70. Ziv, A. & Levin, Z. *J. atmos. Sci.* **31**, 1652–1661 (1974).
71. Szymanski, E. W., Szymanski, S. J., Holmes, C. R. & Moore, C. B. *J. geophys. Res.* **85**, 1951–1953 (1980).
72. Hauser, D. & Amayenc, P. *J. Rech. Atmos.* **14**, 439–455 (1980).
73. Hauser, D. & Amayenc, P. *J. appl. Met.* **20**, 547–555 (1981).
74. Joss, J. & Dyer, R. *Preprints, 15th Conf. on Radar Meteorology*, 179–180 (American Meteorological Society, Boston, 1972).

ARTICLES

Tectonic subsidence, flexure and global changes of sea level

A. B. Watts

Lamont-Doherty Geological Observatory and Department of Geological Sciences, Columbia University, Palisades, New York 10964, USA

Tectonic models for the evolution of passive continental margins predict that following rifting, sediments should progressively onlap basement at the edge of a margin as the lithosphere cools and increases its flexural rigidity with age. The pattern of modelled onlap is strikingly similar to that used by Vail and colleagues to estimate sea-level rise through geological time. This similarity suggests that major portions of stratigraphical sequences at margins may have a tectonic, rather than eustatic, control. The patterns of onlap used by Vail and colleagues may be widespread, however, because several widely separated passive margins rifted at similar times, but they are unlikely to be worldwide.

THE distribution of seas through geological time is the result of two principal factors: global or eustatic changes of sea level and tectonic movements of the Earth's crust[1]. One or both of these factors may cause shorelines to shift and sedimentary facies to change. It may be expected that a relative rise in sea level, due either to a global rise in sea level or to crustal subsidence, would produce a transgressive stratigraphical sequence, while a relative fall of sea level, due either to a global fall in sea level or to crustal uplift, would produce a regressive sequence. The occurrence of a transgressive or regressive sequence at a particular locality in a sedimentary basin, however, is controlled by the rate of sediment influx and the rate and direction of change in the relative position of sea level and the sea floor[2]. It appears that global changes in sea level, due to changes in the volume of land ice or mid-ocean ridge crests, proceeded rapidly enough to explain some of the transgressive and regressive sequences at the edges of epeiric seas during the geological past. But there has been considerable debate[3] on the origin of the transgressive and regressive sequences that occurred during periods of no extensive land ice and relatively constant ridge crest volumes.

Sloss and Speed[4] argued that the major control on stratigraphical sequences in the continental interiors[5] were tectonic movements. They recognized three main episodes or 'modes' during the development of stratigraphical sequences in the continental interiors: 'oscillatory', in which land areas were subject to differential uplift and subsidence; 'emergent', in which land areas were progressively uplifted; and 'submergent', in which land areas progressively subsided. Sloss and Speed[4] noted a similar timing of these modes between widely spaced continents and argued for a synchronous tectonic control over broad regions.

Vail *et al.*[6,7], on the other hand, argued that the major control of sedimentary sequences in continental interiors and margins was global changes of sea level. They recognized several depositional cycles during the development of a stratigraphical sequence, each of which was bounded by surfaces of discontinuity. By identifying these surfaces, using characteristic patterns of seismic reflectors of 'onlap' and 'offlap', they constructed cycle charts at individual localities. Vail *et al.*[6] noted that similar cycles could be recognized at localities in widely separated margins and argued for a eustatic control. They used estimates of the amounts of onlap and offlap to construct a global sea-level curve for the Phanerozoic that was characterized by several short-term fluctuations superimposed on a broad long-term change. Vail *et al.*[7] used the pattern of onlap and offlap to infer sea-level rise and fall respectively. By calibrating their 'global' curve with other estimates of long-term sea-level changes[8,9] they argued that the short-term fluctuations could have magnitudes of up to a few hundred metres[10].

Pitman[9] pointed out that the occurrence of a transgressive or regressive sequence at the edge of a gently sloping shore depends on the rates of change of tectonic movements and global sea level. For example, a transgressive sequence could result from a sea-level fall, if the rate of subsidence of a basin exceeds the rate of sea-level fall. Pitman[9] showed that a major transgression during the Eocene in the North American and African margins could result simply from the interaction of changes in the rates of long-term sea-level fall with the steady, normal subsidence of these margins.

Vail and Todd[11] therefore modified the original statements[7] that coastal onlap and offlap could be directly equated to sea-level rise and fall. Their amended global sea-level curve for the Jurassic, however, was strikingly similar to the original curve, the main difference being the less abrupt sea-level falls.

The nature of the control of stratigraphical sequences has subsequently been widely debated. Hallam[12] argued that a tectonic control on the scale proposed by Sloss and Speed[4] was unlikely, as there was no satisfactory mechanism to explain why widely separated continents would be affected by tectonic movements at similar times. He suggested[13] that the main control on stratigraphical sequences were global changes of sea level and constructed a curve for the Jurassic, similar in overall shape to the Vail *et al.*[6] curve. But as Donovan and Jones[3] have pointed out, without a satisfactory mechanism, the

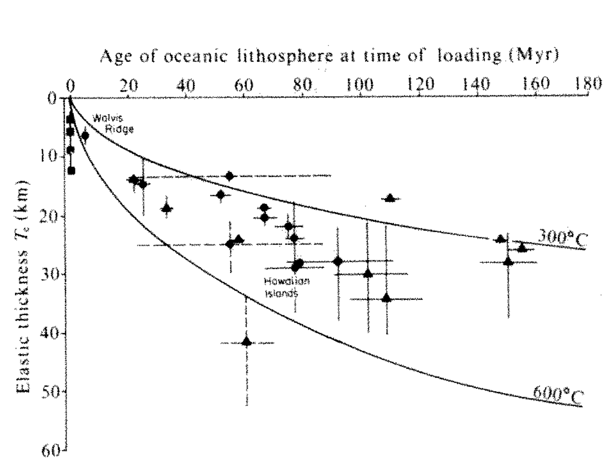


Fig. 1 Plot of elastic thickness T_e against age of the oceanic lithosphere at the time it was loaded[48]. The elastic thickness is determined from observations of the flexural rigidity and assumes Young's modulus = 10^{12} dyn cm^{-2} and Poisson's ratio = 0.25. ●, Seamounts and oceanic islands; ▲, deep-sea trench–outer rise systems; ■, ridge crests; ◆, river delta. Solid lines indicate the 300 °C and 600 °C oceanic isotherms based on the cooling plate model[49].

0028-0836/82/230469—06$01.00

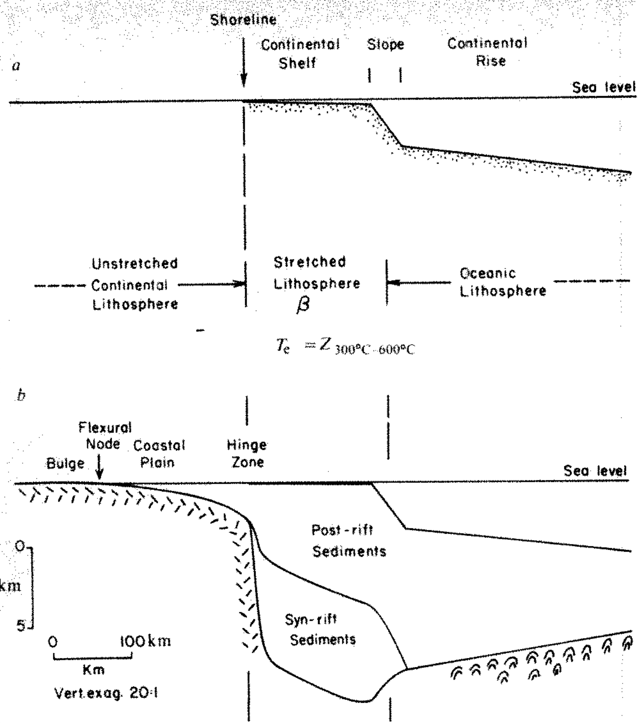


Fig. 2 Thermal and mechanical model for the tectonic evolution of a passive continental margin. *a*, Initial conditions. The tectonic subsidence of the margin is due to thermal contraction following heating and thinning of the lithosphere and crust at the time of rifting. Sediments are assumed rapidly to infill continental shelf, slope and rise regions and to maintain a constant bathymetric profile through time. *b*, Cooling and flexure. The sediments represent a load on the cooling lithosphere which responds by flexure. Sediments that form early in margin history load a relatively weak lithosphere while sediments that form later in its history load a strong lithosphere. The sediments that infill the initial subsidence are referred to as syn-rift while the sediments that infill the thermal subsidence are referred to as post-rift[31]. The model calculations assume an initial lithospheric thickness of 125 km, an initial crustal thickness of 31.2 km, a coefficient of volume expansion of 3.4×10^{-5} °C^{-1}, a mantle temperature of 1,333 °C and initial densities of 2.8 and 3.33 g cm^{-3} for the crust and lithosphere respectively. A uniform density of 2.5 g cm^{3} is assumed for the sediment infill.

large amplitude sea-level changes inferred by Vail *et al.*[6] and Hallam[12] must also remain in doubt. Bally[14] suggested, in fact, that the effects of global sea-level changes are subordinate to tectonics and that the major control on stratigraphical sequences are widespread and correlatable tectonic events associated with major plate reorganizations.

There is little doubt, however, that the sea-level curve of Vail *et al.*[6] represents an important synthesis of high-quality stratigraphical data that apparently has been useful in correlation studies and petroleum exploration. The outstanding question is the interpretation of these curves and whether they represent global sea-level changes or widespread tectonic events. Unfortunately, many of the data used by Vail *et al.*[6] are not generally available so it is difficult to determine whether they have satisfactorily separated eustatic from tectonic effects. This study uses recently developed models for the tectonic evolution of passive margins[15,16] to estimate how tectonics has contributed to stratigraphical sequences.

Tectonic subsidence at passive margins

The knowledge of the tectonic evolution of passive margins has progressed rapidly during the past decade. Backstripping studies[17–21], using data from commercial exploratory and COST-type wells, have shown that the principal factors affecting the post-rift subsidence of passive margins are thermal contraction and sedimentary loading. Factors such as chemical and mechanical compaction, palaeobathymetry, amount of sediment influx and long-term changes in sea level all contribute to the subsidence but their combined effects are small compared with thermal contraction and sediment loading. By correcting the well data for the effects of sedimentary loading it has, therefore, been possible to estimate the form of the thermal contraction and how it may vary across a margin.

Most modelling studies now assume that thermal contraction at passive margins arises from heating and thinning of the crust and lithosphere at the time of intial rifting[22,23]. The main differences between the various models is the manner by which the heating and thinning occurs. Sleep[22] proposed that the thinning was caused by uplift and subaerial erosion while McKenzie[23] proposed that it was caused by a passive, uniform extension of the crust and lithosphere.

There is good observational evidence for crustal extension during the early rifting history of passive margins, particularly at the sediment-starved margins of the eastern Atlantic and portions of the margin off eastern North America. For example, de Charpal *et al.*[24] have mapped listric faults and tilted fault blocks in northern Biscay, and Given[25] and Schlee[26] have mapped graben systems off Nova Scotia and New England. The amount of extension cannot easily be estimated from fault geometry, but seismic refraction data are consistent with a stretching factor[23] of $\beta = 2–3$ over distances of 100–200 km in northern Biscay[27,28].

The stretching model[23] has also been applied to the thickly sedimented margins of the western Mediterranean off southern France[21] and portions of the margin off eastern North America[29,30]. The main difficulty is the lack of information at these margins on the thickness and depositional environment of the sediments formed during rifting and, the tectonic fabric of the underlying basement rocks. For example, the relative proportion of syn-rift to post-rift sediments, an important constraint in the stretching model[31], is not precisely known at these margins. The available evidence[29,30] suggest a stretching factor in the range $\beta = 1.5–6.0$. Unfortunately, there is too little seismic refraction data to constrain these estimates satisfactorily but geoid data are consistent with large amounts of stretching ($\beta = 3–4$) over distances of 100–200 km along portions of these margins (ref. 20 and M. S. Steckler, personal communication).

The model of McKenzie[23] therefore appears to give a general explanation of the available geological and geophysical data at passive margins. Some problems still remain, however, the most important of which are: the discrepancies between estimates of stretching based on fault geometry and deep crustal seismic refraction data[28], the occurrence of seaward-dipping reflectors of possible subaerial volcanic origin at some margins[32] and the relative role of passive and active heating in the thermal evolution of a margin[21].

Flexure

The response of the lithosphere to sediment loads at a passive margin has been modelled using either an Airy or flexure model of isostasy[17,18]. Seismic reflection profiles indicate that active faulting accompanies the early stages of rifting[24–26]. This suggests that an Airy model, in which sediments are locally compensated, is most applicable early in margin evolution. The presence of gently dipping post-rift sediments and a broad coastal plain, however, suggests the flexure model becomes more important later in margin evolution.

The actual lithospheric response to sediment loads is difficult to determine at a passive margin because sediments are continually added to it during long periods of time. The best evidence has come from studies of seamounts and oceanic islands, as they formed during relatively short time periods (1–1.5 Myr). These studies show that the flexural rigidity and the equivalent elastic thickness of oceanic lithosphere, T_e, increases with the age of the lithosphere at the time it was loaded (Fig. 1). Loads formed on young oceanic lithosphere, such as the Walvis ridge, are associated with small values of T_e while loads formed on old oceanic lithosphere, such as the

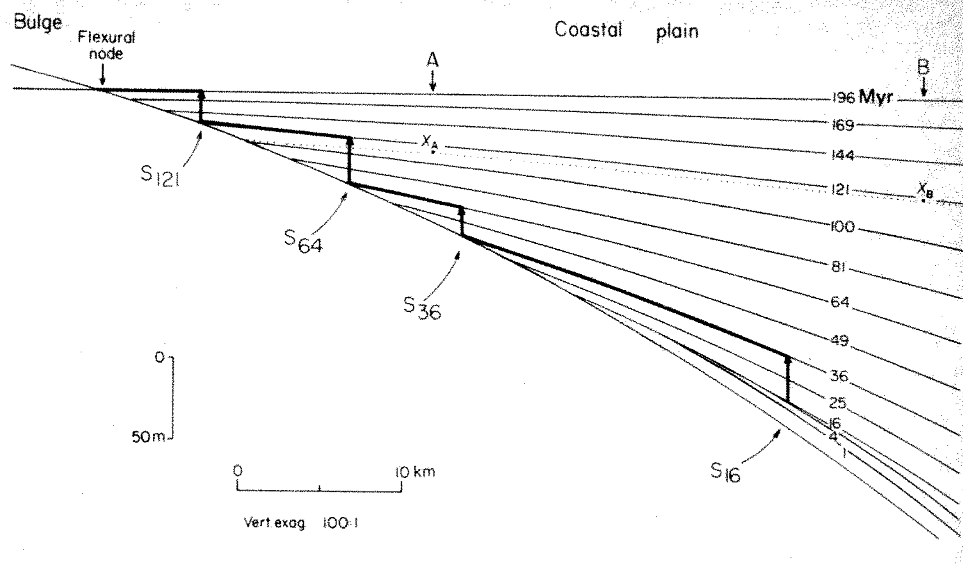


Fig. 3 Coastal plain stratigraphy calculated using the thermal and mechanical model in Fig. 2. The model calculations assume $\beta = 3.0$ and $T_e = Z_{450}$ °C. Solid lines represent individual stratigraphical horizons at different times following rifting. The effects on the stratigraphy of compaction have not been included. The heavy solid lines illustrate the procedure used by Vail *et al.*[7] and this study to estimate the vertical component of coastal onlap. The dotted line indicates the equilibrium shoreline where the rate of change of long-term sea-level fall[9] equals the subsidence rate. Because the coastal plain sediments subside at a slower rate than long-term sea-level rises or falls, there is a tendency, depending on the rate of sediment influx, for transgression below the dotted line (enhancing the effects of flexure) and regression above the line (competing with flexure).

Hawaiian islands, are associated with large values. Figure 1 shows there is good general agreement between T_e and the depth, Z, to the 300–600 °C oceanic isotherms. Thus, flexure studies in the ocean basins suggest that as the oceanic lithosphere increases in age and cools, it becomes more rigid in its response to loads.

Because the lithosphere at passive margins is extensively heated during rifting[22,23], it should therefore become progressively more rigid in its response to sediment loads as it cools following rifting. Sediments formed soon after rifting would be expected to load a relatively hot and weak lithosphere while later sediments would load a relatively cool and strong lithosphere. A flexure model, in which the rigidity of the lithosphere increases with time, explains certain tectonic–stratigraphical features of well sedimented passive margins such as an outer stratigraphical high, a hinge zone and a coastal plain in which younger sediments progressively onlap basement[33].

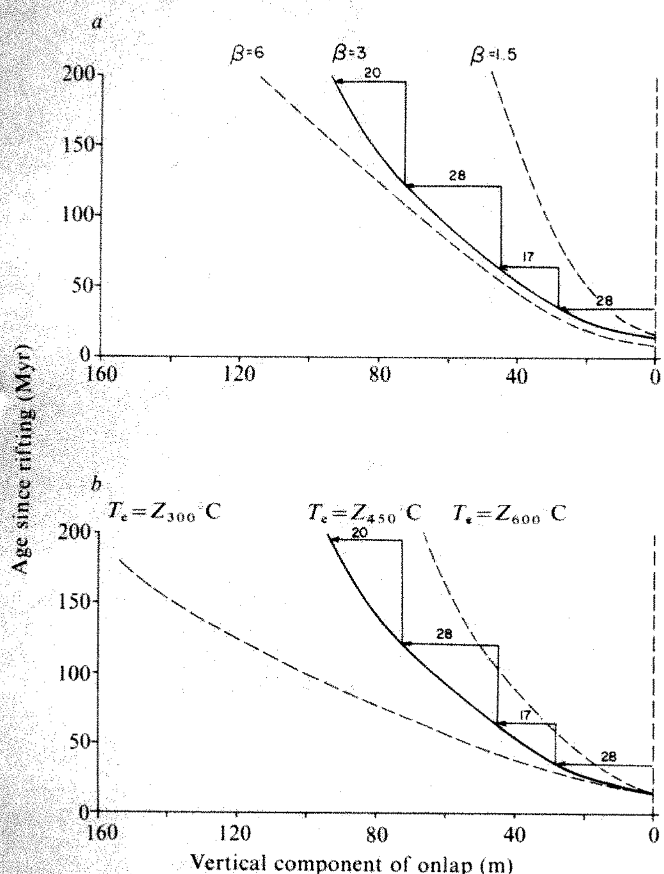


Fig. 4 The vertical component of coastal onlap estimated from the thermal and mechanical model in Figs 2 and 3. Solid lines indicate the individual measured increments of coastal onlap and the heavy solid lines indicate the smoothed change in onlap following rifting. *a*, $T_e = Z_{450}$ °C and β variable. *b*, $\beta = 3.0$ and T_e variable.

Thermal and mechanical models

Thermal and mechanical models have now been constructed[15,16,34] for the stratigraphy of passive margins that combine the effects of thermal contraction and sedimentary loading. In these models, thermal contraction occurs following crustal and lithospheric extension at the time of rifting while sedimentary loading occurs by flexure of a progressively more rigid basement following rifting. Since T_e is a strong function of temperature (Fig. 1), thermal and mechanical effects can be coupled. Thus, T_e can be calculated as a function of time and position following rifting.

Figure 2 shows a simple model for the tectonic evolution of a passive margin that formed by crustal and lithospheric extension at the time of rifting. The region of extension is limited to beneath the shelf and slope and to have a magnitude ($\beta = 3$) and horizontal extent (175 km), similar to present day margins. Sediments are assumed to infill uniformly the cooling margin, maintaining a constant bathymetric profile with time. The model includes the effects of lateral heat conduction across the stretched region and flexure that varies as a function of time and position ($T_e = Z_{450}$ °C).

Figure 3 shows the stratigraphy predicted by the thermal and mechanical model at the edge of the margin. The solid lines indicate stratigraphical horizons at equal increments of the square root of time since rifting. Figure 3 shows that there is initially a significant onlap of sediments onto the basement, due to the abrupt transition in the stretching model from fault-controlled Airy-type subsidence to flexural controlled subsidence. Soon after rifting, the pattern of onlap is abruptly terminated because lateral heat flow from the stretched region to unstretched continental lithosphere causes the coastal plain to remain emergent[15,16]. Beginning about 16 Myr after rifting, however, flexure overcomes the effects of lateral heat flow and younger sediments progressively onlap the basement, due to its increase in flexural rigidity with age.

The model in Fig. 3 is simplified as it assumes an abrupt increase in β across the hinge zone and a sediment influx that

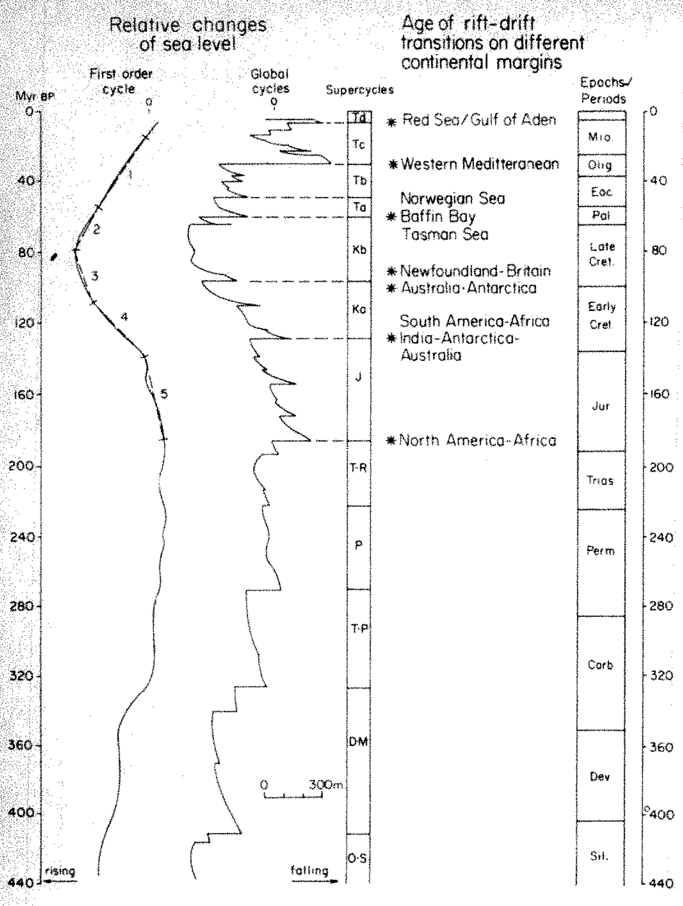


Fig. 5 Global cycles of relative changes of sea level since the Devonian[6] compared with the age of the rift–drift transition at different continental margins[48]. The numbered dashed lines on the 'first-order cycle' are construction lines used to estimate the rates of change of sea-level rise and fall in Fig. 6.

uniformly infills shelf, slope and rise regions. Although there is little information on how β varies across a hinge zone, it seems likely that it would change more smoothly than assumed in the model. An abrupt change in β maximizes the lateral effect of heat flow across the hinge zone so the coastal plain would remain emergent for a longer time following rifting. However, a more gradual transition in β across the hinge zone only slightly decreases the pattern of onlap in Fig. 3. The pattern of onlap would be expected to change if sediments by-passed shelf regions due, for example, to vigorous submarine erosion[26]. A prolonged absence of shelf loading would cause the flexural node (Fig. 2*b*) to migrate seaward towards loaded slope and rise regions, resulting in offlap in the coastal plain. Changes in the pattern of sediment influx seem unlikely, however, to compete with the modelled onlap in Fig. 3 unless it, like flexure, varies systematically during margin evolution.

The model studies therefore suggest that coastal onlap is a characteristic feature of stratigraphical sequences at a cooling margin. This result is important since measurements of the amount of onlap at present day and ancient passive margins is the principal means by which Vail *et al.*[6,7] estimate sea-level changes through time. The models in Figs 2 and 3 have been constructed assuming no sea-level changes through time. Thus, by measuring the amount of onlap in the models we can estimate the contributions of tectonics to stratigraphical sequences.

Vail *et al.*[7] measured onlap by estimating the 'coastal aggradation', or 'vertical component of onlap', in a stratigraphical sequence. Beginning with the lowest point of onlap in a sequence, they successively summed each increment of onlap until the highest point of onlap is reached. Figure 3 illustrates, on the modelled stratigraphy, the procedure used by Vail *et al.*[7]. Starting with S_{16}, the lowest point of onlap in the sequence, the first increment of onlap is 28 m and the time interval is 16–36 Myr. Successive increments of 17, 28 and 20 m give a total amount of onlap for the sequence of 93 m. Figure 4 shows the largest amount of onlap (100–160 m) occurs for the greatest amount of stretching and weakest plate, while the smallest amount (45–65 m) occurs for the least stretching and strongest plate. The form of the onlap is similar in each model, generally following the shape of an exponential curve, with an initial rapid increase followed by a more gentle increase.

Although they did not detail their method, Vail *et al.*[7] used coastal onlap to estimate sea-level rise through time. They used a modal average of three or more correlative cycles of onlap from different continents to construct a global sea-level curve for the Phanerozoic. But, as Fig. 4 shows, coastal onlap is a characteristic feature of the tectonic evolution of a cooling passive margin and does not require sea-level changes to produce it. Therefore, if the model predictions are correct, some correlation should exist between the beginning of major cycles of onlap in the Vail *et al.*[6] curve and the age of the rift–drift transition in passive margins, formed as a result of continental break-up.

Figure 5 shows there is an excellent visual correlation between the beginning of the supercycles of Vail *et al.*[6] and the age of the rift–drift transition at margins formed by the breakup of the Pangea supercontinent. There is a striking similarity between the form of the modelled onlap and individual supercycles (Fig. 5). This similarity suggests, therefore, that major portions of the Vail *et al.*[6] curve may have a tectonic, rather than eustatic, control.

Vail *et al.*[6], however, terminate their supercycles by rapid sea-level falls (Fig. 5). The tectonic models predict regressive sequences early in the development of a margin (Fig. 3) but, they do not predict them at higher levels in a stratigraphical sequence. Therefore, if the magnitudes of the sea-level falls deduced by Vail *et al.*[6] are correct then, factors other than flexure must be invoked to explain them.

There is some doubt that Vail *et al.*[6] have correctly estimated the amount of sea-level fall from stratigraphical sequences[14,35]. The main problem is that they determine sea-level fall by measuring the vertical component of onlap between the highest point of onlap in an underlying sequence directly to the lowest point of onlap in an overlying sequence. Because the record is largely removed by erosion during a sea-level fall, they cannot use the 'incremental method' that was used to measure onlap. Thus, the effects of tectonics during the sea-level fall are not satisfactorily accounted for.

There is good evidence[36], however, for major periods of regression in the geological record, although their origin is in some doubt. For example, Grasty[37] has related major regressive events to crustal thickening associated with orogeny while Ager[36] has related them to long-term changes in global sea level, due to changes in mid-ocean ridge crest volumes.

The effects of long-term changes in sea level[9,6] on the modelled stratigraphy are illustrated in Fig. 6. The open triangles (Fig. 6) indicate the age since rifting when the rate of sea-level fall equals the rate of model subsidence and the shoreline would stabilize. For small times following rifting the seas would rise faster than the model subsides, causing a transgression, while for longer times the seas would fall faster, causing a regression. The actual effect of long-term sea-level changes depends, however, on the relative rates of sediment influx and subsidence[2]. For the modelled stratigraphy in Fig. 3, a transgressive sequence would be expected to develop soon after rifting, enhancing the pattern of onlap produced by flexure, while for later times a regressive sequence would develop, competing with flexure. The timing of the resulting seaward shift in onlap would depend on position, but for the model in Fig. 3 it occurs about 112–117 Myr following rifting.

The model predictions are difficult to compare with the Vail *et al.*[6] curve as this curve represents the summation of several shifts in the pattern of onlap from different positions in widely

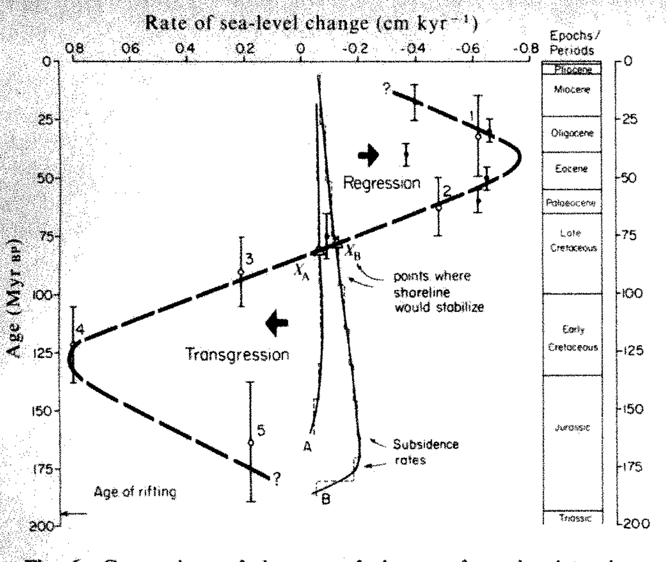


Fig. 6 Comparison of the rate of change of sea level to the subsidence rate at two points (A, B) in the thermal and mechanical model in Fig. 3. The rate of change of sea level is based on the long-term sea-level curve of Pitman[9] and the first-order cycle of Vail *et al.*[6]. ●, Based on Table 2 of ref. 9; ○, a straight line fit to the first-order cycle of Vail *et al.*[6] (Fig. 5). △, The time in Myr BP that the rate of sea-level fall equals the rate of subsidence and the shoreline, in the absence of other effects, would be in equilibrium.

spaced margins. There is evidence though, from published geological cross-sections of the US Atlantic coastal plain[38], of a progressive onlap of Jurassic to early Cretaceous sediments onto basement that is terminated by a seaward shift in the overall pattern of onlap about 120 Myr after rifting. Thus, the model in Fig. 3 seems to agree with the overall stratigraphy of this relatively old margin.

The tectonic models are still unable, however, to explain the widespread rapid short-term falls of sea level in the Vail *et al.*[6] curve. Thus, without a satisfactory mechanism, the origin of these falls remains in doubt.

Discussion

The model studies suggest that tectonics is an important control on the development of stratigraphical sequences at passive margins. Flexure produces patterns of onlap, remarkably similar in form to those used by Vail *et al.*[7] to infer sea-level rise. Although Vail *et al.*[7] attempted to remove tectonic effects by summing increments of coastal onlap in a sequence, they did not correct for flexural effects that vary as a function of time and position. They considered only a limited role for tectonics that was generally equivalent to assuming sedimentary loading occurs by flexure of a lithosphere of uniform flexural rigidity.

As has already been pointed out, however, flexure cannot explain all the occurrences of onlap in stratigraphical sequences; neither can it explain the occurrence of offlap at high levels in stratigraphical sequences. The Cenomanian and Callovian transgressive sequences[36] extended too far into the interiors of the continents to be explained by flexure, so other factors, such as long-term changes in global sea level due to changes in mid-ocean ridge crests, are required to explain them. Similarly, the Oligocene regressive sequences of the eastern US[35], western Australia[39], and the northern North Sea[6] cannot be explained by flexure. These relatively old margins were subsiding at relatively slow rates during the Oligocene so the relatively rapid long-term fall in global sea level[9] since 85 Myr BP may have contributed to these sequences. Furthermore, flexure cannot explain transgressive sequences such as the Hettangian[40] of south-west Britain, which were too far north to have been affected by the Tethys passive margin[41], or regressive sequences such as the Oligocene of southern Africa[42]. Tectonics, but not necessarily flexure, may have contributed to these sequences[40,42].

Figure 7 shows the overall pattern of stratigraphical sequences predicted by the tectonic models for different age margins. Following rifting each margin would be expected to show its own pattern of onlap as it cools and becomes progressively more rigid with age. But as each margin increases in age, there is a greater tendency to be affected by long-term changes in sea level. For example, the fall in long-term sea level from 85 to 15 Myr BP (ref. 9) would cause a shift in the pattern of onlap at approximately similar times in each margin. The patterns of onlap would also be affected, of course, by the rise in sea level before 85 Myr, but this would only enhance the effect of flexure. Figure 7 shows that although there is a similarity in the timing of the shift in onlap at each margin, the overall pattern of onlap is strikingly different between each margin. The models predict, in fact, that there would be little correlation in patterns of onlap between margins unless they rifted at similar times. Thus patterns of onlap, of the type used by Vail *et al.*[7] may be widespread, because many widely spaced margins rifted at similar times, but they are unlikely to be worldwide.

Unfortunately, there is presently too little stratigraphical data available to test the model predictions in Fig. 7. Vail *et al.*[7] only published a few sea-level curves from separate localities and most of these extend for the duration of only one supercycle. However, Todd and Mitchum[43] published sea-level curves for Texas Gulf Coast and West Africa; localities associated with the rifting of North America from Africa at ~196 Myr BP (ref. 44). Although it is not possible to compare their curves directly with the model predictions, both curves show a steady increase in the amount of onlap during the Jurassic. The main difference is that the models do not predict the short-term falls in sea level, the largest being in the Valanginian (125–130 Myr BP). As Todd and Mitchum[43] pointed out, however, a restricted Valanginian sequence can only be inferred from seismic profiles in the Gulf Coast.

This article has focused on the nature of the control on stratigraphical sequences in present day passive margins but tectonics, in the form of lithospheric flexure, may also have been a major control in the continental interiors. For example, Vail *et al.*[6] constructed a sea-level curve for the Devonian and Cambrian; presumably based on data from the Michigan basin and the western and eastern US orogenic belts. They show a

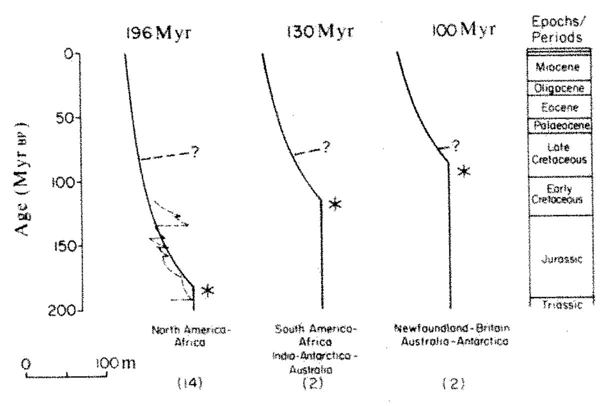


Fig. 7 The pattern of onlap predicted for passive margins of different ages. The heavy solid line represents the pattern of onlap predicted from the thermal and mechanical models with $\beta = 3$ and $T_e = Z_{450}$ °C. The heavy dashed line indicates the shift in the pattern of onlap caused by a fall in long-term sea level since 85 Myr BP. The actual shift in the pattern would vary across a margin. The fine dashed line indicates the sea-level curve for the Texas Gulf Coast based on Todd and Mitchum[44]. This curve has not been calibrated[44] and is shown for comparison only. There is a significant difference in the patterns of onlap between each margin because they rifted at separate times. The numbers in parentheses indicate the number of localities that Vail *et al.*[6] apparently used in each margin. Because they used a modal average of patterns of onlap from individual localities, the pattern of onlap from the North America–Africa margin should dominate their global sea-level curve.

prominent pattern of onlap for each of these periods which they interpreted as a sea-level rise. The increase in width of the Michigan basin during Devonian[45] can be explained[46], however, by an increase in flexural rigidity of the lithosphere with age, and does not require sea-level changes to produce it. Furthermore, the apparent progressive onlap of sediments onto the western and eastern margins of the North American craton during the Cambrian[47] can be explained by flexure, although the large horizontal extent of the Albertan may require long-term sea-level changes in addition.

This is not to imply that a completely satisfactory model exists yet for the tectonic evolution of sedimentary basins in present day and ancient passive margins. There is too little seismic reflection and refraction data at present day margins that satisfactorily constrains the deep crustal structure implied by the thermal models. The occurrence of coastal onlap at a passive margin, however, is the consequence of the cooling of the lithosphere following rifting and does not depend on a particular type of thermal model. The actual separation of tectonic effects from stratigraphical sequences remains a difficult problem that will require careful studies of seismic, geological and palaeontological data from localities in widely separated margins. The problem is important for future consideration, however, as it has major implications for tectonics, lithospheric mechanics, correlative stratigraphy and the thermal history of the Earth.

I thank G. Karner, W. C. Pitman and M. S. Steckler for special assistance. This work has been supported by NSF grants OCE 79-26308 and EAR 81-09473, and the Arthur D. Storke Memorial Fund. Lamont-Doherty contribution no. 3324.

Received 15 January; accepted 2 April 1982.

1. Ager, D. V. *The Nature of the Stratigraphic Record* (Macmillan, London, 1973).
2. Grabau, A. W. *Principles of Stratigraphy* Vol. 2 (Dover, New York, 1960).
3. Donovan, D. T. & Jones, E. J. W. *J. geol. Soc. Lond.* **136**, 187–192 (1979).
4. Sloss, L. L. & Speed, R. C. *Spec. Publ. Soc. Econ. Palaeont. Miner.* **22**, 98–119 (1974).
5. Sloss, L. L. *Bull. geol. Soc. Am.* **74**, 93–114 (1963).
6. Vail, P. R., Mitchum, R. M. & Thompson, S. *Am. Ass. petrol. Geol. Mem.* **26**, 83–97 (1977).
7. Vail, P. R., Mitchum, R. M. & Thompson, S. *Am. Ass. petrol. Geol. Mem.* **26**, 63–81 (1977).
8. Hays, J. D. & Pitman, W. C. *Nature* **246**, 18–22 (1973).
9. Pitman, W. C. *Bull. geol. Soc. Am.* **89**, 1389–1403 (1978).
10. Vail, P. R. & Hardenbohl, J. *Oceanus* **22**, 71–79 (1979).
11. Vail, P. R. & Todd, R. G. in *Petroleum Geology of the Continental Shelf of North-West Europe* (eds Illing, L. V. & Hobson, G. D.) 216–235 (Heyden, London, 1981).
12. Hallam, A. *Palaeogr. Palaeoclimatol. Palaoecol.* **23**, 1–32 (1978).
13. Hallam, A. *J. geol. Soc.* **138**, 735–743 (1981).
14. Bally, A. W. *Am. geophys. Un. Geodyn. Ser.* **1**, 5–20 (1981).
15. Steckler, M. S. thesis, Columbia Univ. (1981).
16. Steckler, M. S. & Watts, A. B. *Am. geophys. Un. Geodyn. Ser.* **8**, 184–196 (1981).
17. Watts, A. B. & Ryan, W. B. F. *Tectonophysics* **36**, 24–44 (1976).
18. Steckler, M. S. & Watts, A. B. *Earth planet Sci. Lett.* **41**, 1–13 (1978).
19. Keen, C. E. *Can. J. Earth Sci.* **16**, 505–522 (1979).
20. Watts, A. B. & Steckler, M. S. *Maurice Ewing Symp. Ser.* **3**, 218–234 (1979).
21. Steckler, M. S. & Watts, A. B. *Nature* **287**, 425–429 (1980).
22. Sleep, N. H. *Geophys. J. R. astr. Soc.* **24**, 325–350 (1971).
23. McKenzie, D. P. *Earth planet. Sci. Lett.* **40**, 25–32 (1978).
24. de Charpal, O., Guennoc, P., Montadert, L. & Roberts, D. G. *Nature* **275**, 706–711 (1978).
25. Given, M. M. *Bull. Can. Petrol. Geol.* **25**, 63–91 (1977).
26. Schlee, J. S. *Bull. Am. Ass. petrol. Geol.* **65**, 26–53 (1981).
27. LePichon, X. & Sibuet, J.-C. *J. geophys. Res.* **86**, 3708–3720 (1981).
28. Avedik, F. *et al. Proc. R. Soc.* (in the press).
29. Royden, L. & Keen, C. E. *Earth planet Sci. Lett.* **51**, 343–361 (1980).
30. Watts, A. B. *Am. Ass. petrol. Geol. Educ. Course Note Ser.* **19**, (1981).
31. Sclater, J. G. & Christie, P. A. *J. geophys. Res.* **85**, 3711–3739 (1980).
32. Hinz, K. *Geol. Jb. Reihe E. Geophysik. Heft* **22**, 3–28 (1981).
33. Watts, A. B. & Steckler, M. S. *Oceanol. Acta* **4**, 143–153 (1981).
34. Beaumont, C., Keen, C. E. & Boutilier, R. *Geophys. J. R. astr. Soc.* (in the press).
35. Olsson, R. K., Miller, K. G. & Ongrady, T. E. *Geology* **8**, 549–554 (1980).
36. Ager, D. V. *J. geol. Soc. Lond.* **138**, 159–166 (1981).
37. Grasty, R. L. *Nature* **216**, 779–780 (1967).
38. Brown, P. M., Miller, J. A. & Swain, F. M. *U.S. geol. Surv. Pap.* **76** (1972).
39. Quilty, P. G. *Geology* **5**, 336–340 (1977).
40. Donovan, D. T., Horton, A. & Ivimey-Cook, H. C. *J. geol. Soc. Lond.* **136**, 165–173 (1979).
41. Bernouille, D. *Am. Ass. petrol. Geol. Educ. Course Note Ser.* **19** (1981).
42. Siesser, W. G. & Dingle, R. V. *J. Geol.* **89**, 83–96 (1981).
43. Todd, R. G. & Mitchum, R. M. *Am. Ass. petrol. Geol. Mem.* **26**, 145–163 (1977).
44. Van Houten, F. B. *Bull. Am. Ass. petrol. Geol.* **61**, 79–99 (1977).
45. Sloss, L. L. & Scherer, W. *Geol. Soc. Am. Mem.* **142**, 71–8 (1975).
46. Haxby, W. F., Turcotte, D. L. & Bird, J. M. *Tectonophysics* **36**, 57–75 (1976).
47. Kay, M. & Colbert, E. H. *Stratigraphy and Life History* (Wiley, New York, 1965).
48. Watts, A. B., Karner, G. D. & Steckler, M. S. *Proc. R. Soc.* (in the press).
49. Parsons, B. & Sclater, J. G. *J. geophys. Res.* **82**, 803–827 (1977).

Human EJ bladder carcinoma oncogene is homologue of Harvey sarcoma virus *ras* gene

Luis F. Parada, Clifford J. Tabin, Chiaho Shih & Robert A. Weinberg

Center for Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Examination of homologies between retroviral oncogenes and transforming sequences defined by transfection reveals that the human bladder carcinoma (EJ) oncogene is homologous to the Harvey sarcoma virus oncogene (ras). *Structural analysis limits the region of homology to a 3.0-kilobase* SacI *fragment of the EJ oncogene. Both EJ and* ras *DNA probes detect similar transcripts in transfectants derived from bladder carcinoma cell lines.*

TWO groups of cellular oncogenes have been discovered during the past decade. The first consists of genes that were characterized by virtue of their association with retroviruses. The prototype of this class is the *src* gene of avian sarcoma virus. Several experiments have indicated that this gene was acquired from the chicken genome by an avian retrovirus, and has been exploited by the chimaeric virus to transform cells[1,2]. Once incorporated into the viral genome, expression of the *src* gene is driven by viral controlling elements and is no longer responsive to control mechanisms that governed its expression while in the cellular chromosome. In addition to the *src* gene, this group includes at least 12 other gene sequences, each associated with a different chimaeric retrovirus[3,4]. These genes are conserved over great evolutionary distances[2,5] implying that they mediate essential cellular or organismic functions.

A second class of cellular transforming genes has been detected by the experimental route of DNA transfection. Recent reports have indicated that the DNAs of some non-virally induced tumour cell lines can induce transformation when applied to mouse fibroblast monolayers. These tumour cell lines are derived from chemically induced animal tumours[6–9] and from human tumours of spontaneous origin[7,10–12].

The two classes of oncogenes have many properties in common, the most striking of which is the apparent origin of both types of genes from normally benign, cellular genetic elements. Because of this and other parallels, we undertook a search to determine whether the two groups of genes shared any members in common. Such overlap would have far-reaching consequences for our understanding of the mechanisms of viral and non-viral carcinogenesis.

0028-0836/82/230474—05$01.00

Table 1 Oncogene probes

Oncogene‡	Related virus or tumour	Clone	Sequence complexity	Source	Ref.
v-*abl*	Murine Abelson leukaemia virus	pAblsub9	3.0 kbp	J. Y. J. Wang and D. Baltimore	*
v-*erb*	Avian erythroblastosis virus	PAE-*Pvu*II	2.5 kbp	T. Gonda and J. M. Bishop	28
v-ST-*fes*	Snyder–Theilen feline sarcoma virus	PST-3	0.6 kbp	C. Sherr	29
c-Ha-*ras*1 (rat)	Murine Harvey sarcoma virus	LHXB-3	2.3 kbp	R. Ellis and E. Scolnick	30
v-Ha-*ras*	Murine Harvey sarcoma virus	BS9	0.45 kbp	R. Ellis and E. Scolnick	21
c-*mos* (Human)	Murine Moloney sarcoma virus	pHM1	2.75 kbp	G. Vande Woude	†
v-*myc*	Avian myelocytomatosis virus	puMyC3-*Pst*	1.5 kbp	T. Gonda and J. M. Bishop	31
v-*src*	Avian sarcoma virus	SRA-2	0.8 kbp	T. Gonda and J. M. Bishop	32
	Human EJ bladder carcinoma	pEJ6.6	6.6 kbp	C. Shih	18

* J. Y. J. Wang and D. Baltimore, in preparation. † G. Vande Woude, personal communication. ‡ v- indicates viral gene; c- indicates cellular gene.

Detection of homologies by nucleic acid hybridization

The search for relatedness between the two groups of genes depended on detection of nucleic acid sequence homologies between individual members of each group. One group consisted of a series of seven retrovirus-associated *onc* genes known to be unrelated or only distantly related to one another[4]. The other group studied was a collection of seven tumour oncogenes that had been defined by transfection[7,10]. Nucleic acid sequence probes have been derived for the retrovirus-associated genes, whereas only a few of the transfection-derived genes have been isolated in the form of molecular clones. We therefore used the virus-derived *onc* probes to survey the DNAs of cells which had acquired, via transfection, copies of the second class of genes. The survey was performed using the Southern gel–filter transfer procedure[13]. Virus-derived *onc* probes were from several sources (see Table 1).

We used each retrovirus-related *onc* probe to search for novel cross-reacting fragments in each of several transfectants. For example, an analysis using the v-*abl* probe is shown in Fig. 1: lane *a* represents the endogenous mouse sequences detected in the DNA of untransfected NIH 3T3 cells by the v-*abl* sequence probe. These 'background' bands of 28, 10 and 6 kilobases (kb) were also found in the DNAs of all transfectants (Fig. 1, lanes *b*, *d–g* and *i*). Figure 1*c* shows detection of the human homologue of the v-*abl* probe; lanes *d–g* show an analysis of DNAs of mouse cells transfected with four different human oncogenes. None of the DNAs contains any fragments beyond those present in the untransfected mouse control (lane *a*). We concluded that none of these transfected cells acquired the human homologue of the v-*abl* sequence. From a similar analysis in lanes *h* and *i*, we concluded that the rabbit bladder carcinoma oncogene is also not related to the *abl* gene. A more equivocal interpretation came from analysis of the DNA of a mouse cell transfected with a mouse fibroblast oncogene (Fig. 1*a*, *b*). Due to the lack of species-specific fragment markers, we were unable to rule out the identity of the *abl* and fibroblast oncogene. However, knowledge of their restriction enzyme cleavage sites excludes identity[14,15].

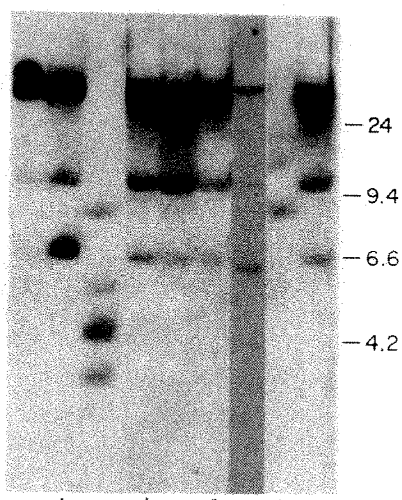


Fig. 1 Southern blot analysis of digested cellular DNAs from various transfectants probed with v-*abl* specific probe (see Table 1). 10 μg of each DNA were digested with endonuclease *Eco*RI, fractionated by electrophoresis through a 1% agarose gel and transferred to nitrocellulose paper[13]. The filters were incubated with 5×10^6 c.p.m. of nick-translated[39] ^{32}P-labelled Abelson virus specific probe. The DNAs analysed were from the following cell lines: *a*, NIH 3T3; *b*, Y5-1; *c*, HeLa; *d*, A5-2; *e*, SH-1-1; *f*, SW-2-1; *g*, EJ-6-1; *h*, rabbit embryo fibroblast; *i*, RBC-1. Lanes *a–f* and *h*, *i* are from one filter; *g* is from a different filter. The transfected cell lines are described in Table 2. Migration of *Hin*dIII-digested λ DNA fragments are shown at the right (in kilobase pairs, kbp).

Figure 2 shows a further Southern blot analysis of DNAs that were prepared from several cell lines derived by transfection of NIH 3T3 cells with DNAs of human tumour cell lines. The probe used was the rat cellular homologue of Harvey sarcoma virus, c-Ha-*ras*1, given by R. Ellis and E. Scolnick[16,17]. A novel fragment of 9.5 kb was found in the DNA of a mouse cell transfected with DNA of the EJ human bladder carcinoma cell line (Fig. 2*c*). In contrast, no novel fragments were present in the transfectants derived from other tumour cell lines (Fig. 2*d–g*).

Of the seven *onc* probes used in this survey, only one detected the presence of novel DNA fragments in transfection-derived cell lines. The results of this comparative oncogene survey are summarized in Table 2. Below we consider the relationship between the c-Ha-*ras*1 gene and the human bladder carcinoma gene suggested by Fig. 2.

Homology of the two oncogene DNAs

A series of tests was performed to further substantiate the relationship between the EJ bladder oncogene and the c-*ras* oncogene. Figure 3*A* shows that nine cell lines derived by transfection of EJ bladder carcinoma DNA have all acquired novel DNA fragments reactive with the c-Ha-*ras*1 probe. DNA from untransfected mouse cells (Fig. 3*A*, lane *a*) does not exhibit any of these novel fragments. The *Bam*HI-digested DNAs of the various transfectants exhibit differently sized novel fragments because of rearrangements occurring during the transfection process. Note that the oncogene of the T24 human bladder carcinoma is closely related to that of the EJ bladder carcinoma[18,19]. Figure 3*A*, lane *m*, indicates that a mouse transfectant carrying the T24 oncogene also contains in its DNA a novel acquired fragment reactive with the probe.

To further define the linkage between the EJ oncogene and the c-Ha-*ras*1 homologous sequences, we analysed the DNAs of three EJ transfectants derived by transfer of *Bam*HI-cleaved DNA (Fig. 3*B*). The donors of these DNAs were secondary transfectants derived previously by two serial passages of the bladder carcinoma oncogene. As endonuclease *Bam*HI does

Table 2 Comparison of transfected oncogenes with retroviral *onc* probes

	Transfected cell lines							
onc Probes	Y5-1-1	EJ-6-1	A5-2	SH-1-1	SW-2-1	HL60-1-9	B104-1-1	RBC-1
v-*abl*	*	–	–	–	–	–	NT	–
v-*erb*	*	–	–	–	–	–	NT	NT
v-*fes*	*	–	–	–	–	–	–	NT
c-Ha-*ras*1 (rat)	*	+	–	–	–	–	–	NT
c-*mos* (Human)	*	–	NT	NT	–	–	NT	NT
v-*myc*	*	–	–	–	–	–	–	NT
v-*src*	*	–	–	NT	–	–	NT	–

The oncogene probes shown are described in Table 1. The transfected cell lines listed are of the following origins: Y5-1-1 is derived from two serial passages of the oncogenic DNA from a 3-methylcholanthrene-induced mouse fibroblast cell line, MCA-16 (refs 6, 33); EJ-6-1 is a secondary transfected cell line derived from DNA of a human bladder carcinoma cell line, EJ[7]; A5-2 is a primary transfected cell line derived from human lung carcinoma cell line, A549 (ref. 34); SH-1-1 is a secondary transfected cell line derived from DNA from human neuroblastoma cell line, SK-N-SH (J. Føgh, personal communication); SW-2-1 is a secondary transfected cell line derived from human colon carcinoma cell line, SW-480 (refs 10, 35); HL60-1-9 is a secondary transfected cell line derived from human leukaemia cell line, HL60 (refs 10, 36); B104-1-1 is a secondary transfected cell line derived from rat neuroblastoma cell line, B104 (refs 7, 37); RBC-1 is a primary transfected cell line derived from rabbit bladder carcinoma cell line, RBC[7,38]. + Indicates hybridization of the probe to novel bands in addition to hybridization with NIH 3T3 cellular sequences; – indicates a well controlled negative correlation. * Apparent negative correlation that is not based on species-specific DNA fragment sizes that distinguish donor DNAs from resident chromosomal DNAs of the recipient cells. NT, not tested.

not inactivate the EJ oncogene[18], the transfection of *Bam*HI-cleaved secondary DNA ensured that almost the only human fragment present in resulting tertiary transfectants was the 6.6-kb *Bam*HI fragment bearing the EJ oncogene. The three DNAs of the tertiary transfectants were analysed after *Eco*RI cleavage as it was thought they might have lost some *Bam*HI sites during the second transfection. The DNAs of the three transfectants (Fig. 3*B*, lanes *a–c*) all showed acquired fragments reactive with the c-Ha-*ras*1 probe. This demonstrated that the linkage between the EJ oncogene and the c-Ha-*ras*1 homologous sequences could not be broken by *Bam*HI cleavage.

A further comparison between the genes depended on the fact that the EJ bladder carcinoma oncogene was one that we have recently isolated as a molecular clone[18]. The EJ human bladder oncogene has been cloned as a biologically active *Eco*RI fragment of 16 kb carried by a Charon 4A λ phage vector and termed Φ631. A biologically active 6.6-kb *Bam*HI fragment subclone has been inserted into plasmid vector pBR322 and termed pEJ6.6. All endonucleases shown to cleave within this 6.6 kb insert (Fig. 4) inactivate the focus-inducing activity of this DNA (ref. 18 and C. Shih, unpublished results).

Using the EJ 6.6-kb *Bam*HI fragment and the c-Ha-ras1 oncogene clone as sequence probes, we analysed the DNA fragments homologous to these genes in normal human DNA. Both probes detected a 6.6-kb *Bam*HI fragment (Fig. 3*A*, *C*, lanes *b*) and a 23-kb *Eco*RI fragment (Fig. 3*B*, *D*, lanes *b*) in

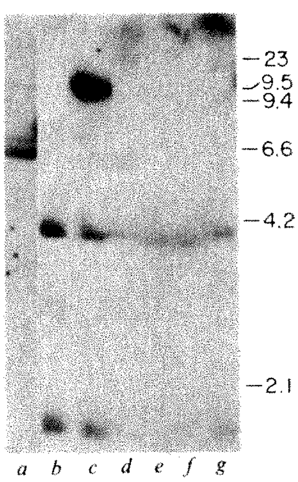


Fig. 2 Southern blot analysis of DNA from transfected cells with rat cellular *ras* probe (c-Ha-*ras*1). ^{32}P-labelled c-Ha-*ras*1 DNA was prepared as described in Fig. 1 legend and incubated with a filter carrying *Bam*HI-digested DNAs from the following cell lines: *a*, HeLa; *b*, NIH 3T3; *c*, EJ-6-1; *d*, A5-2; *e*, SH-1-1; *f*, SW-2-1; *g*, HL-60-9. Lanes *b–g* are from the same filter. The cell lines are described in Table 2.

human DNA (see also ref. 20). Furthermore, the EJ bladder oncogene probe detected the same novel fragments in transfected mouse lines (Fig. 3*C*, lanes *c–k*) that were previously detected using the c-Ha-*ras*1 probe (Fig. 3*A*, lanes *c–k*).

Endonuclease-cleaved Φ631 DNA was immobilized on a cellulose nitrate filter and probed with the c-Ha-*ras*1 sequences. Figure 4*a* indicates that homology between c-Ha-*ras*1 and the EJ clone is limited to the 6.6-kb *Bam*HI fragment of the bladder carcinoma oncogene; lane *b* further reduces the domain of homology between the two oncogenes to a 3.0-kb *Sac*I fragment within the 6.6-kb *Bam*HI fragment. Figure 4*f–i* shows a similar experiment to that of lanes *b–e* but in this case BS-9, a v-Ha-*ras* probe, was used (provided by Drs D. Lowy and E. Scolnick). BS-9 is a *ras* specific subclone of the Harvey sarcoma virus genome[21] and is ~450 base pairs (bp) long. This probe includes the 5′ half of the v-*ras* gene. Comparison of left and right panels of Fig. 4 confirms that the viral probe and c-Ha-*ras*1 cross-hybridize with identical fragments of the EJ oncogene DNA. Double digests (Fig. 4*g*, *i*) with *Sac*I+*Kpn*I or *Sac*I+*Xba*I indicate that the v-Ha-*ras* homology straddles the *Kpn*I and *Xba*I cleavage sites indicated at the top of Fig. 4. Most of the reactivity of the c-Ha-*ras*1 probe lies in the larger of the two fragments created by these digests (Fig. 4*c*, *e*). The deduced alignments between the EJ oncogene and the v-Ha-*ras* and c-Ha-*ras*1 probes are shown at the top of Fig. 4. The direction of transcription, deduced from the results of the present study and previous data[16,17], is from right to left on the map.

Taken together, these results indicate that the EJ bladder oncogene is closely linked to the human homologue of the rat c-Ha-*ras*1 sequences. Although the limits of the c-Ha-*ras*1 structural sequences have been well defined[16,17,20], the corresponding sequences of the EJ gene have not yet been mapped. Thus, the data above cannot exclude the possibility that the two genetic elements were adjacent to one another rather than congruent.

Analysis of transcripts homologous to the clones

The transcripts encoded by these genes were analysed to further establish their relationship to one another. We examined the RNAs of transfected cells for molecules reactive with the two oncogene probes. As shown in Fig. 5, the two probes each detected transcripts of 1.2 and 5.1 kb in both the parental tumour cell line and in EJ- and T24-transfected mouse cell lines (see also ref. 22). These transcripts were not detected in untransfected NIH 3T3 cells. Thus, introduction of the EJ oncogene into mouse cells results in synthesis of RNAs that are homologous with the rat c-Ha-*ras*1 gene. As discussed below, these data support a congruency between the functionally active region of the EJ gene and that of the c-*ras* gene.

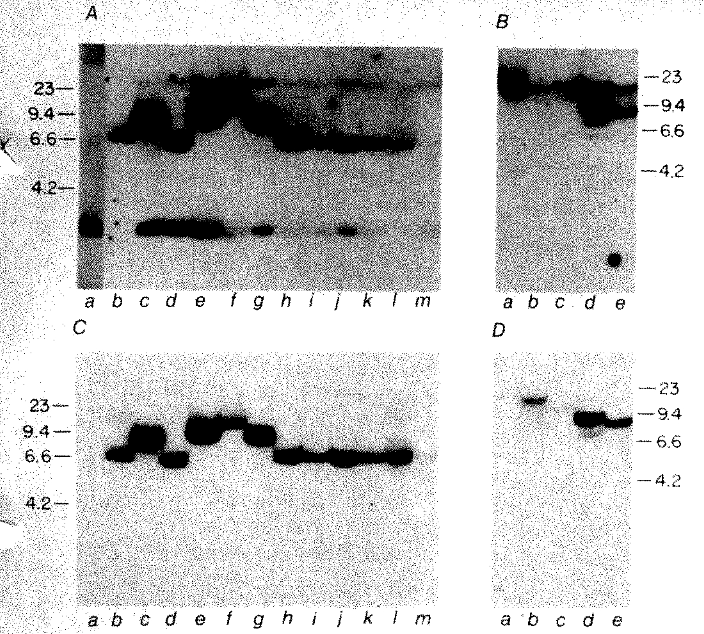


Fig. 3 Analysis of DNAs from EJ and T24 transfectants using the c-Ha-*ras*1 and EJ oncogene probes. DNAs were digested with endonuclease *Bam*HI (*A*, *C*) or *Eco*RI (*B*, *D*) and analysed as described in Fig. 1 legend. The filters shown *A* and *B* were incubated with 5×10^6 c.p.m. of c-Ha-*ras*1 DNA probe (2×10^8 c.p.m. μg^{-1}) and exposed for autoradiography for 16 h. The adsorbed ^{32}P-labelled probe was removed from the filters by washing in 0.1 M NaOH, 0.5 M NaCl, 1.0 mM EDTA. The washed filters were exposed to film for 48 h to verify that all the radioactive signal had been removed (not shown). The filters were then incubated with 5×10^6 c.p.m. of nick-translated pEJ6.6 DNA (6×10^7 c.p.m. μg^{-1}). *C* and *D* represent autoradiography after 16 h. The DNAs analysed in panels *A* and *C* were from the following cell lines: *a*, NIH 3T3; *b*, HeLa; *c*, EJ-6-1; *d*, EJ-2-R1; *e*, EJ-2-R5; *f*, E-4-R4-B; *g*, EJ-6-1; *h*, EJ-6-2-R; *i*, EJ-6-3; *j*, EJ-1-2; *k*, EJ-3-2; *l*, T-24 human cells[19]; *m*, T24-8-5. *B* and *D* display DNAs from: *a*, NIH cells; *b*, HeLa cells; *c*, EJ-6-2 (*Bam*)-1 cells; *d*, EJ-6-2(*Bam*)-2 cells; *e*, EJ-4(*Bam*)-1 cells. In *A* and *C*, lanes *c–f* and *h–k* represent DNA from cell lines that were derived from independent serial transfections of the EJ tumour cell oncogene; lanes *g* and *c* are duplicates. Lane 1 contains DNA from human bladder carcinoma cell line T24. Lane *m* contains DNA from a secondary transfectant of T24. Lanes *c–e* of *B* and *D* show an analysis of DNAs from tertiary transfectants induced by exposure to *Bam*HI-cleaved secondary transfectant DNA.

Discussion

The present data strongly suggest an evolutionary homology between the EJ human bladder carcinoma oncogene and the rat c-Ha-*ras*1 gene. We now consider the experimental basis for this conclusion and its implications.

The relatedness between the EJ and c-Ha-*ras*1 genes was first noted when we demonstrated that a cell acquiring the EJ oncogene also carried a novel DNA fragment reactive with the c-Ha-*ras*1 probe. As the linkage between the c-Ha-*ras*1 homologous sequence and the EJ oncogene was not broken by endonuclease *Bam*HI, we concluded that the two genetic elements lay within the same 6.6-kb *Bam*HI-generated fragment. These data alone were consistent with the two elements being either physically adjacent or congruent with one another. To resolve this ambiguity, we analysed the transcripts encoded by the two oncogenes. The EJ parental tumour and its derived transfectants, all express 5.1- and 1.2-kb transcripts that react with the oncogene probe. The fact that these transcripts are also present in cells transfected with *Bam*HI-cleaved DNA (data not shown) implies that the entire transcriptional unit of the 5.1-kb RNA is found within the confines of the 6.6-kb *Bam*HI-generated fragment (Fig. 3*B*, *D*).

It could be argued that the transcripts detected in transfected mouse cells are of murine origin. In this case, their synthesis would be induced indirectly in the mouse cells by the acquired human oncogene rather than being encoded directly by the human gene. We consider this unlikely, as the rat c-Ha-*ras*1 probe used had a three-fold higher specific radioactivity than the human EJ probe, but yielded a 3–4-fold lower signal intensity on autoradiography (Fig. 5). This must reflect the relatively lower affinity of the rat probe for human transcripts present in the EJ transfectants. Thus, we conclude that the two probes both detect RNAs transcribed largely, if not entirely, from the human EJ oncogene template.

A similar, if not identical, pair of transcripts has been found by colleagues working with the c-Ha-*ras*1 gene and its RNAs[22]. As both the EJ and c-Ha-*ras*1 homologous 5.1-kb transcriptional units lie within the 6.6-kb *Bam*HI fragment, we conclude that the two genes are congruent with one another rather than adjacent. Consistent with this structural homology is a functional analogy in that both genes are able to induce fibroblast transformation.

The *ras* genes encode proteins of molecular weight 21,000 (ref. 23). Immunoprecipitation of metabolically labelled lysates of transfected cells has detected a protein of this size (R. Finkelstein and R.A.W., unpublished observations). However, the association of this protein with the bladder oncogene will only be well established after its detailed peptide structure has been analysed. Knowledge of the function of this protein may elucidate the important steps in human bladder carcinogenesis.

This p21 polypeptide represents a strong candidate for the protein mediating transformation of a human tumour cell. It is

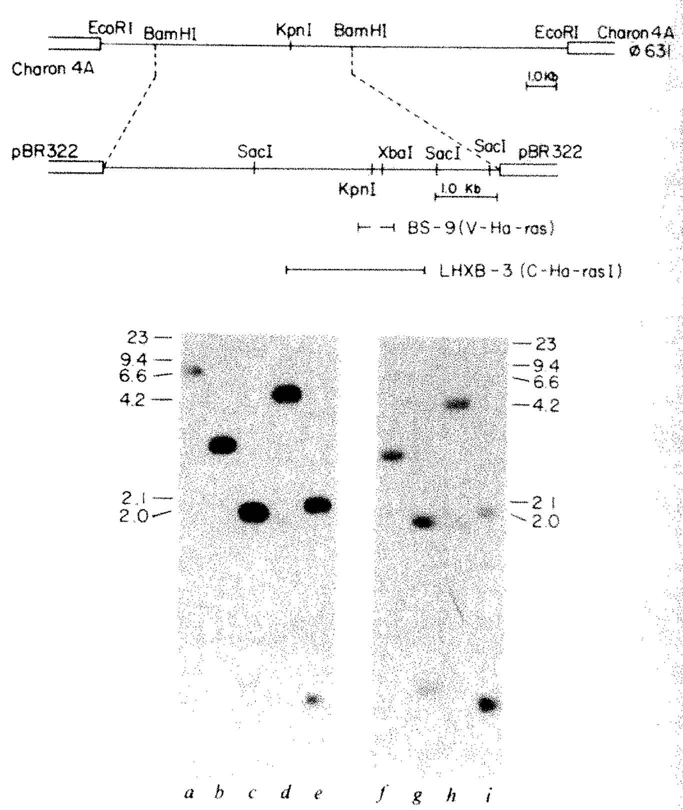


Fig. 4 Alignment of c-Ha-*ras*1 and of v-Ha-*ras* (BS-9) with the physical map of the EJ oncogene. DNA from the EJ–Charon 4A clone Φ631 was digested with several restriction enzymes. DNA (0.5 μg) was loaded onto each lane before electrophoresis and blot transfer. Identical nitrocellulose filters were prepared and incubated with ^{32}P-labelled DNA of c-Ha-*ras*1 (lanes *a–e*) or with HaSV subclone BS-9 (lanes *f–i*). Φ631 DNA was cleaved with *Bam*HI (*a*); *Bam*HI + *Sac*I (*b*, *f*); *Sac*I + *Kpn*I (*c*, *g*); *Bam*HI + *Xba*I (*d*, *h*); *Sac*I + *Xba*I (*e*, *i*). The alignment shown is accurate to within 200 nucleotides.

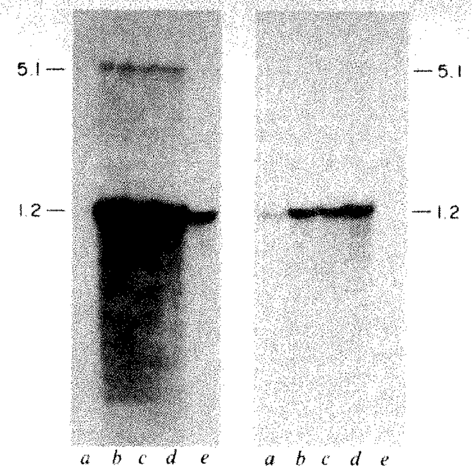


Fig. 5 Cellular polyadenylated RNAs analysed using pEJ6.6 and c-Ha-*ras*1 ^{32}P-labelled probes. In the left-hand panel, nick-translated c-Ha-*ras*1 DNA (2×10^{8} c.p.m. μg^{-1}) was used to probe RNA isolated from the following cell lines: *a*, NIH 3T3; *b*, T24-8-1 (a secondary transfectant derived from T24 human cell line); *c*, EJ-4(R1)-2; *d*, EJ-6-2; *e*, EJ human bladder. In the right-hand panel, nick-translated pEJ6.6 (6.6×10^{7} c.p.m. μg^{-1}) was used to probe the following cell lines:*a*, EJ; *b*, EJ-6-2; *c*, EJ-4(R1)-2; *d*, T24-8-1; *e*, NIH 3T3. The polyadenylated RNAs were prepared by the technique of Varmus *et al.*[40]. The RNA was then fractionated by electrophoresis through formaldehyde-containing gels and transferred to nitrocellulose (B. Seed and D. Goldberg, in preparation). ^{32}P-labelled probes, prepared as described in Fig. 1 legend, were annealed to the immobilized RNAs[41]. Bands which represent sequences homologous to the probes were visualized by autoradiography. Molecular weights were determined by comparison with markers obtained from *in vitro* run-off transcription of the adenovirus late promoter[42] and are shown in kilobases.

one of the first proteins implicated directly in the oncogenic conversion of a cell following its transformation by non-viral agents, and has been localized at the inner surface of the plasma membrane in Harvey sarcoma virus (HaSV)-transformed cells[24]. If this localization applies also to the EJ bladder carcinoma cells, then the transforming protein of these cells, the p21, should not display extracellular antigenic determinants. In this case, any tumour-specific surface antigens displayed by the bladder carcinoma cell should be encoded by genetic elements other than the oncogene itself.

The present work has several other implications. Perhaps the most apparent is that a single proto-oncogene can be activated by different molecular processes. The c-Ha-*ras*1 oncogene of the rat became activated via its affiliation with retrovirus sequences, forming the chimaeric Harvey sarcoma virus[21,25]. As demonstrated in the accompanying article[26], the human c-Ha-*ras*1 gene is also capable of oncogenic activation after it becomes linked *in vitro* to retrovirus promoter sequences. The mode of activation of the EJ oncogene is different, but not yet understood. Presently evidence suggests that the EJ oncogene and its normal human allelic counterpart sequence are indistinguishable by restriction enzyme site mapping[18]. It is possible that its activation depends on minor structural alterations, such as point mutations.

The relatedness between the EJ oncogene and that of a transforming retrovirus represents an advance in our understanding of the molecular basis of human bladder carcinoma. This stems from the fact that the structure and function of the *ras* genes and their gene products have been extensively studied[16–18,20–23].

We have been unable to demonstrate other homologies between retrovirus *onc* genes and the transfection-derived tumour genes, but this may merely reflect the small repertoire of tumour genes presently available in cloned form. As other genes become available for study, additional connections will probably be found.

Two paradoxes seem to be raised by the unexpected association of a rat sarcoma oncogene with a human bladder oncogene. First, this work implies the ability of the c-Ha-*ras*1 gene to act in unrelated tissue environments. The rat gene, when carried in Harvey sarcoma virus, can induce sarcomas and erythroleukaemias[27] while its human counterpart is now implicated in the genesis of bladder carcinomas[10–12,18,19] We consider it possible that the *ras* oncogene of either species is capable of transforming a wide range of target tissues, only a small portion of which has been studied experimentally.

Second, we have suggested that the precursor of the EJ oncogene, now identified as c-Ha-*ras*1, represents a preferred target for activation during bladder carcinogenesis[18]. It seems unlikely that the bladder urothelium was the site of acquisition of the *ras* gene during the events that let to the creation of the chimaeric HaSV genome. The two routes of oncogene activation must involve different molecular mechanisms which probably occur at different sites in the organism. Each mode of activation may be favoured by different predisposing factors present in different tissues. For example, in the bladder, the c-Ha-*ras*1 gene may be in a configuration particularly susceptible to mutational activation whereas in certain other tissues it may be expressed in a manner favouring the recombinational events that lead to creation of chimaeric retroviruses.

We thank our colleagues for providing the *onc* probes used in this study. This research was supported by US National Cancer Institute grants CA17537 and CA26717 to R.A.W.

Received 8 April; accepted 17 May 1982.

1. Spector, D. H., Varmus, H. E. & Bishop, J. M. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4102–4106 (1978).
2. Stehelin, D., Varmus, H. E., Bishop, J. M. & Vogt, P. K. *Nature* **260**, 170–173 (1976).
3. Klein, G. (ed.) *Advances in Viral Oncology: Cell Derived Oncogenes* (Raven, New York, 1981).
4. Coffin, J. M. *et al. J. Virol.* **40**, 953–957 (1981).
5. Shilo, B-Z. & Weinberg, R. A. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6789–6792 (1981).
6. Shih, C., Shilo, B-Z., Goldfarb, M., Dannenberg, A. & Weinberg, R. A. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5714–5718 (1979).
7. Shih, C., Padhy, L. C., Murray, M. & Weinberg, R. A. *Nature* **290**, 261–264 (1981).
8. Hopkins, N., Besmer, P., DeLeo, A. B. & Law, L. W. *Proc. natn. Acad. Sci. U.S.A.* **78**, 7555–7559 (1981).
9. Lane, M. A., Sainten, A. & Cooper, G. M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 5185–5189 (1981).
10. Murray, M. J. *et al. Cell* **25**, 355–361 (1981).
11. Krontiris, T. G. & Cooper, G. M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1181–1184 (1981).
12. Perucho, M. *et al. Cell* **27**, 467–476 (1981).
13. Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).
14. Shilo, B-Z. & Weinberg, R. A. *Nature* **289**, 607–609 (1981).
15. Goff, S. P., Gilboa, E., Witte, O. N. & Baltimore, D. *Cell* **22**, 777–785 (1980).
16. DeFeo, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 3328–3332 (1981).
17. Ellis, R. W. *et al. Nature* **292**, 506–511 (1981).
18. Shih, C. & Weinberg, R. A. *Cell* (in the press).
19. Goldfarb, M., Shimizu, K., Perucho, M. & Wigler, M. *Nature* **296**, 404–409 (1982).
20. Chang, E. H., Gonda, M. A., Ellis, R. W., Scolnick, E. M. & Lowy, D. R. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
21. Ellis, R. W. *et al. J. Virol.* **36**, 408–420 (1980).
22. Ellis, R. W., DeFeo, D., Furth, M. & Scolnick, E. M. *Cell* (in the press).
23. Shih, T. Y., Weeks, M. O., Young, H. A. & Scolnick, E. M. *Virology* **96**, 64–79 (1979).
24. Willingham, M. C., Pastan, I., Shih, T. Y. & Scolnick, E. M. *Cell* **19**, 1005–1014 (1980).
25. Harvey, J. J. *Nature* **204**, 1104–1105 (1964).
26. Chang, E. H., Furth, M. E., Scolnick, E. M. & Lowy, D. R. *Nature* **297**, 479–483 (1982).
27. Chesterman, F. C., Harvey, J. J., Dourmashkin, R. R. & Salaman, M. H. *Cancer Res.* **26**, 1759–1768 (1966).
28. Vennstrom, B., Fanshiev, L., Moscovici, C. & Bishop, J. M. *J. Virol.* **36**, 575–585 (1980).
29. Sherr, C. J., Fedele, L. A., Oskarsson, M., Maizel, J. & Vande Woude, G. *J. Virol.* **34**, 200–212 (1980).
30. Chang, E. H., Gonda, M. A., Ellis, R. A., Scolnick, E. M. & Lowy, D. R. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
31. Vennstrom, B., Moscovici, C., Goodman, H. & Bishop, J. M. *J. Virol.* **39**, 625–631 (1981).
32. DeLorbe, W. J., Luciw, P. A., Goodman, H. M., Varmus, H. E. & Bishop, J. M. *J. Virol.* **36**, 50–61 (1980).
33. Rapp, V. R., Nowinski, R. C., Reznikoff, C. A. & Heidelberger, C. *Virology* **65**, 329–409 (1975).
34. Giard, D. J. *et al. J. natn. Cancer Inst.* **51**, 1417–1421 (1973).
35. Leibovitz, A. *et al. Cancer Res.* **36**, 4562–4569 (1976).
36. Collins, E. J., Gallo, R. C. & Gallagher, R. E. *Nature* **270**, 347–349 (1977).
37. Schubert, D. *et al. Nature* **249**, 224–226 (1974).
38. Summerhayes, I. C. & Franks, L. M. *J. natn. Cancer Inst.* **62**, 1017–1021 (1979).
39. Rigby, P. W., Dieckmann, M., Rhodes, C. & Berg, P. *J. molec. Biol.* **13**, 237–251 (1977).
40. Varmus, H. E., Quintrell, N. & Ortiz, S. *Cell* **25**, 23–36 (1981).
41. Wahl, G. M., Stern, M. & Stark, G. R. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3683–3687 (1979).
42. Manley, J. L., Fire, A., Cano, A., Sharp, P. A. & Gefter, M. L. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3855–3859 (1980).

Tumorigenic transformation of mammalian cells induced by a normal human gene homologous to the oncogene of Harvey murine sarcoma virus

Esther H. Chang*†, Mark E. Furth‡, Edward M. Scolnick‡ & Douglas R. Lowy*

* Dermatology Branch and ‡ Laboratory of Tumor Virus Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205, USA

A normal human gene homologous to the p21 ras *oncogene of Harvey murine sarcoma virus induced oncogenic transformation and high p21* ras *levels in murine fibroblasts when this gene was ligated to a control element (the long terminal repeat) from a murine or feline retrovirus. These results indicate that high levels of a gene product encoded by a normal human oncogene can induce tumorigenic transformation.*

TRANSFORMING retroviruses have been isolated from diverse sources, including birds, rodents and non-human primates (for review see ref. 1). The *in vivo* tumorigenicity of these viruses has been correlated with their ability to induce focal transformation of tissue culture cells. These infected cells are fully transformed by the criteria of anchorage-independent growth *in vitro* and tumorigenicity in animals. As viral DNA genomes and subgenomic viral DNA fragments can induce similar transformation in mouse NIH 3T3 fibroblasts[2,3], this assay has been used to define the viral sequences required for focal transformation. Two components are essential: the viral transforming oncogene (v-*onc*) and the viral long terminal repeat (LTR), a control element which permits transcription of the v-*onc* gene[4,5].

More than ten v-*onc* genes have been described, each of which has been found to be transduced from a closely related, evolutionarily conserved, cellular oncogene (collectively called c-*onc*; reviewed in ref. 6). Using v-*onc* DNAs as probes and antibodies directed against v-*onc*-specified proteins, RNA and protein expression of some c-*onc* genes has been detected in normal cells. However, the level of v-*onc* expression in virally transformed cells is usually higher than that of c-*onc* expression in normal cells, suggesting that elevated levels of c-*onc* gene expression might induce oncogenic transformation. This prediction has recently been proven at least for two rodent c-*onc* genes: c-*mos*[7,8], the mouse sequences from which v-*mos* of Moloney murine sarcoma virus (Mo-MuSV) was derived, and c-Ha-*ras*1 (ref. 9), a rat gene closely related to the transforming gene (v-Ha-*ras*) of the rat-derived Harvey (Ha) MuSV. The gene product of v-Ha-*ras* is a protein (p21) of molecular weight 21,000 (p21), and low levels of an immunologically cross-reacting non-viral p21 have been found in normal cells from various vertebrate species[10]. While v-Ha-*ras* contains no intervening sequences (IVS) within the p21 coding sequences, rat c-Ha-*ras*1 has three IVS between its four exons. In contrast to most other c-*onc* genes, mouse c-*mos* does not contain intervening sequences, and transcription of this gene in normal cells has not been observed[8,11]. Nonetheless, both rat c-Ha-*ras*1 and mouse c-*mos* DNAs were capable of inducing cellular transformation of mouse NIH 3T3 cells, provided that a viral LTR was ligated upstream from the cellular *onc* gene. Transformation was associated with high RNA levels of c-Ha-*ras* or c-*mos*, respectively, and in the case of c-Ha-*ras*, high levels of p21.

† Present address: Department of Pathology, Uniformed Services University of the Health Sciences, Bethesda, Maryland 20014, USA.

Given the evolutionary conservation of c-Ha-*ras*1 (refs 12, 13), we have sought to determine whether its counterpart in normal human cells has oncogenic potential similar to that of rat c-Ha-*ras*1. If the rat and human genes behaved similarly, it would suggest that the capacity of this oncogene to induce transformation was a property of this gene in any species. As part of our studies of *ras* genes, we have recently detected and cloned from normal human DNA four genes which share homology with v-*ras*[13]. The *ras* sequence in one of these human genes (called human c-Ha-*ras*1) was closely related to v-Ha-*ras* and had a structure very similar to that of rat c-Ha-*ras*1. We have now found that this human gene can induce oncogenic transformation of NIH 3T3 cells when an LTR is present upstream. The transformants form tumours in nude mice and

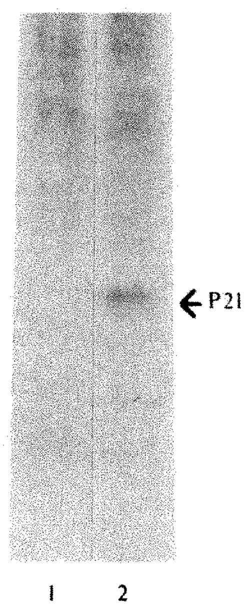


Fig. 1 Immunoprecipitation of endogenous p21 from normal human cells. WI-38 human fibroblasts were metabolically labelled with ^{35}S-methionine for 18 h and lysates were prepared and immunoprecipitated (5×10^6 c.p.m. of TCA-precipitable counts) by a monoclonal antibody as described elsewhere[15,35]. Dissolved immunoprecipitates were then electrophoresed on a 10.5% SDS-polyacrylamide gel. Lane 1, anti-α-tubulin antibody; lane 2, v-Ha-*ras* p21-specific antibody (Y13-238). This monoclonal antibody was also used in Figs 4, 6.

0028-0836/82/230479—05$01.00

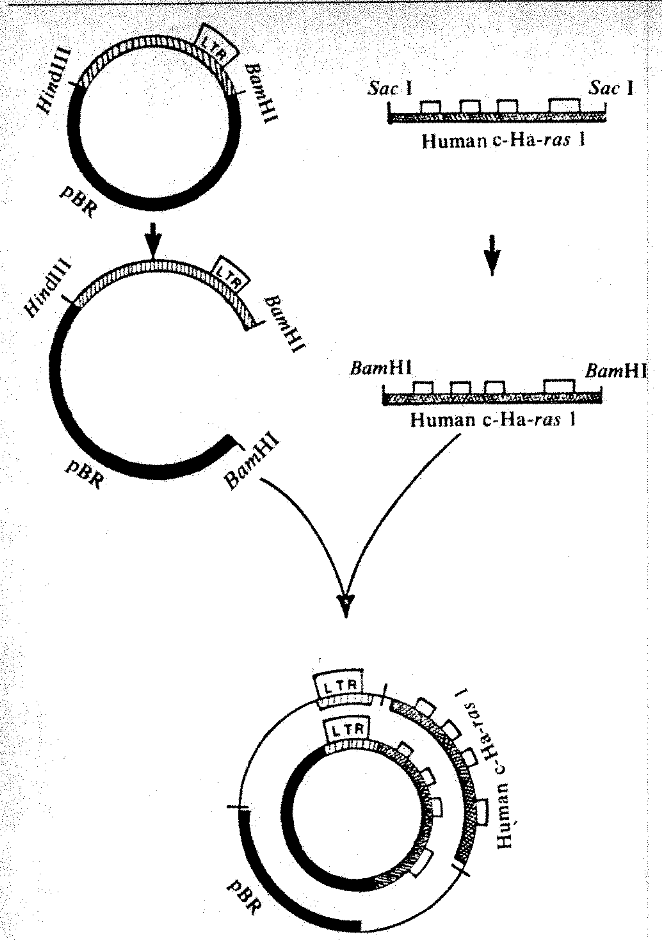


Fig. 2 Schematic representation of the construction of the Ha-MuSV LTR–human c-Ha-*ras1* recombinant. To ligate an LTR upstream from the human c-Ha-*ras1* gene, we constructed a chimaeric recombinant plasmid between the 2.9-kb *SacI* fragment of human c-Ha-*ras1* and a pBR322 clone which already contained the Ha-MuSV LTR. This *SacI* fragment contains all the c-Ha-*ras1* sequences which hybridized to v-Ha-*ras*[12]. The pBR322 LTR clone chosen for this construction (clone P-14) contains a 2.5-kb *HindIII/BamHI* fragment of Ha-MuSV DNA; these sequences, which are derived from a circularly permuted molecule of Ha-MuSV[18], span the region from a *HindIII* site in the 3' end of the Ha-MuSV DNA to the LTR and include 0.3 kb of 5' sequences beyond it to a *BamHI* site; they do not include v-Ha-*ras* sequences. To facilitate insertion of the c-Ha-*ras1 SacI* fragment at the unique *BamHI* site in clone P-14, an attempt was made, as shown in the figure, to add *BamHI* oligonucleotide linkers (Collaborative Research) to the *SacI* ends, after the termini were made flush with DNA polymerase I (Klenow). This change in the ends of the 2.9-kb c-Ha-*ras1* fragment was, however, apparently not achieved because when DNAs from two clones (6-2 and 6-9) which hybridized to a Ha-*ras* probe were analysed, they contained deletions around the inserted sites as shown in the inner circular DNA at the bottom of the figure, instead of the predicted *BamHI* insert shown in the outer circular DNA. There was also a deletion in the 3' sequence of the Ha-MuSV DNA upstream from the LTR. The deletions were 1.4 kb in the 3' sequences of Ha-MuSV DNA, 0.3 kb at the junction between the Ha-MuSV and human c-Ha-*ras1* sequences, and 1.8 kb at the junction between pBR322 and human c-Ha-*ras1*; 1.7 kb of this latter deletion was contained in pBR322 sequences. Despite these deletions, the viral LTR and the region of c-Ha-*ras1* homologous to v-Ha-*ras* were both intact and joined in a single cloned molecule. pBR322 DNA is represented by solid shading, Ha-MuSV DNA by diagonal shading, and human c-Ha-*ras1* DNA by cross-hatching. The open rectangles in c-Ha-*ras1*, which is shown left-to-right in the 5'–3' orientation, represent the exons in the gene.

contain high levels of a p21 which is similar to that specified by the rat gene.

Normal human cells contain low levels of p21

We have recently reported that p21 *ras* is a gene family which consists of at least two prototypic gene types[14]: those closely related to *ras* of Ha-MuSV (Ha-*ras*) and those closely related to *ras* of Kirsten (Ki) MuSV. Ki-MuSV is a rat-derived virus whose v-*ras* encodes a p21 which is immunologically related to the Harvey p21. Of the four human *ras* genes we have detected[13], human c-Ha-*ras1* contains three IVS, as does rat c-Ha-*ras1*, and it is closely related to v-Ha-*ras*. Two of the other human *ras* genes are more closely related to v-Ki-*ras*, and one is distantly related to v-Ha-*ras*.

Because p21 *ras* is a gene family consisting of at least Harvey and Kirsten type genes, it is not clear which of these genes encode the low p21 levels detected in normal cells. The p21 in rodent fibroblasts have recently been found to be predominantly of the Kirsten type[15]. This analysis was made possible by the isolation of a panel of eight monoclonal antibodies to v-Ha-*ras* p21 (ref. 15), the specificities of which have been defined by the viral forms of p21: some recognize Harvey viral p21 but not Kirsten viral p21, whereas others detect both the Harvey and Kirsten forms of p21. In rodents, the type specificity of these monoclonal antibodies apparently also extends to p21 species encoded by Harvey and Kirsten type c-*ras* genes[15,16].

To determine whether human cells contain Harvey type p21, as defined above by the monoclonal antibodies, ^{35}S-methionine-labelled extracts of normal human diploid fibroblasts (WI-38 cells) were immunoprecipitated with a v-Ha-*ras* p21-specific monoclonal antibody. In contrast to rodent fibroblasts[15], a distinct 21K band was detected with this antibody (Fig. 1). This result suggests that these human cells contain significant levels of Ha-*ras* p21.

LTR activation of human c-Ha-*ras1* transforming activity

The two previously cloned c-*onc* retroviral homologues, mouse c-*mos* and rat c-Ha-*ras1*, which have been shown to possess oncogenic potential, were both inactive by themselves when assayed for their ability to transform NIH 3T3 cells[7–9]. To demonstrate their capacity to induce morphological transformation, it was necessary to ligate a viral LTR upstream from the

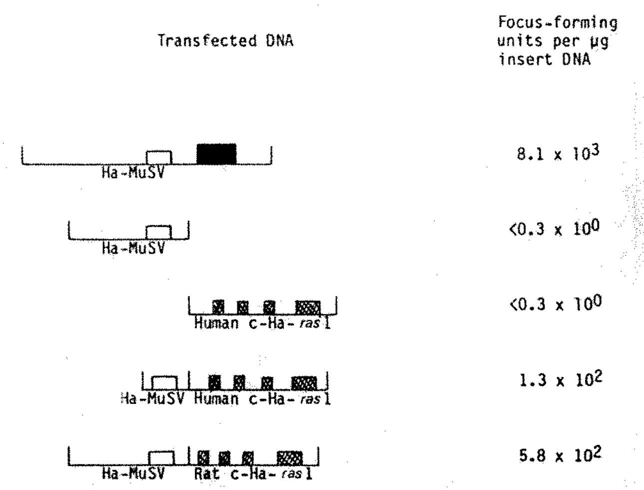


Fig. 3 Transformation induced by the normal Harvey *ras* oncogene. DNAs were precipitated with calcium chloride[36] and 0.2 ml was added to 35-mm dishes seeded on the previous day with 2.25×10^5 NIH 3T3 cells. Calf thymus DNA (25 µg ml^{-1}) was used as carrier. Each DNA insert shown was cloned in pBR322; 0.1 or 0.2 µg of insert DNA was used per dish. The DNAs were added without digestion by a restriction endonuclease. The cells were treated as described elsewhere[2], except that DM50 was not used. Foci were counted 16 days later. The viral LTR is represented by an open rectangle, v-Ha-*ras* by a solid rectangle, and the exons of c-Ha-*ras1* (defined by their homology to v-Ha-*ras*) by cross-hatched rectangles. Ha-MuSV DNA was genomic clone H-1, which is a 5.4-kb circularly permuted copy of the viral DNA genome[5,37]; Ha-MuSV LTR DNA was clone P-14, described in Fig. 2 legend; human c-Ha-*ras1* DNA contained the 2.9-kb *SacI* fragment also described in Fig. 2 legend. In other experiments, a 6.6-kb *BamHI* fragment of c-Ha-*ras1* which contained this 2.9-kb *SacI* fragment also failed to induce foci. Ha-MuSV LTR–human c-Ha-*ras1* DNA was clone 6-2 (clone 6-9 gave similar results in other experiments) described in Fig. 2 legend. Ha-MuSV LTR–rat c-Ha-*ras1* DNA was clone P-14 into which a 2.3-kb fragment of rat c-Ha-*ras1* had been inserted (by M. Haiken, R. Ellis and E.H.C.) at the *BamHI* site. This 2.3-kb fragment of rat c-Ha-*ras1* extends from *EcoRI* to *XbaI* in clone LHBE[9]; these sites were converted (by R. Ellis) to *BamHI* by adding oligonucleotide linkers and cloned at the *BamHI* site in pBR322 (this latter clone is called LXBE).

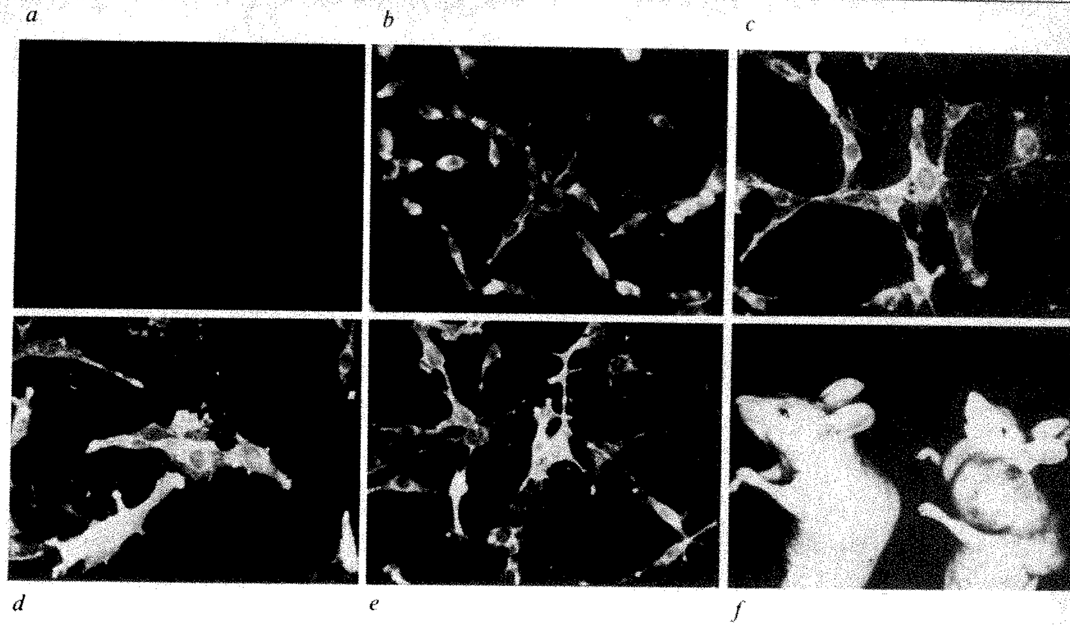


Fig. 4 Detection by immunofluorescent staining of p21 in NIH 3T3 cells transformed by Harvey v-*ras* and c-*ras* genes; tumour formation of human c-*ras* transformants. *a–e*, NIH 3T3 cells and derivatives transformed by Harvey v-*ras* or c-*ras* genes were tested for expression of p21 by indirect immunofluorescent staining with a monoclonal antibody specific for Harvey *ras* p21: *a*, NIH 3T3 cells; *b*, clonal line transformed by Ha-MuSV (viral non-producer Ha821); *c*, clonal line transformed by rat c-Ha-*ras*1 ligated to LTR of Ha-MuSV 9; *d*, focus of cells transformed by human c-Ha-*ras*1 ligated to LTR of Ha-MuSV (clone 6-2); *e*, focus of cells transformed by human c-Ha-*ras*1 ligated to LTR of FeSV. Trypsinized cells were grown on 8-spot printed glass slides (Roboz Surgical) pretreated with 0.1% gelatin, for 18 h at 37 °C in Dulbecco's modified essential medium (DMEM)+10% fetal calf serum. After washing with phosphate-buffered saline (PBS), the cells were fixed by incubation at 25 °C for 10 min in 3.7% formaldehyde in PBS, followed by incubation in absolute methanol at −10 °C for 4 min. The slides were rinsed briefly in acetone (−10 °C) and air-dried. The fixed cells were incubated at 37 °C for 45 min with monoclonal antibody Y13-238 (ref. 15). The antibody was concentrated and partially purified from culture supernatants of cells grown in serum-free medium, and was used at a final concentration of ~20 μg ml^{-1} in PBS with 0.4% (w/v) bovine serum albumin (BSA). After washing with PBS, the cells were incubated at 37 °C for 45 min with rhodamine-labelled antibodies against rat IgG, prepared in goats (Cappel Laboratories). The fluorescent antibody was affinity-purified on a column of rat IgG by Dr Jurgen Wheland, NCI, and was used at a concentration of 125 μg ml^{-1} in PBS with 0.4% BSA. The slides were washed extensively with PBS, mounted under coverslips (no. 1) with Elvanol (DuPont), and observed using a Zeiss microscope equipped with epifluorescent optics. The cells were photographed under UV light using a ×40 Planapochromat objective. Photographs were matched for times of exposure and development, except that the development time for *a* was reduced to enhance the low level of background staining; at matched development this panel would be completely black. Controls of normal and transformed cells stained after incubation with normal rat serum or irrelevant monoclonal antibodies are indistinguishable from *a* (data not shown). Final magnification is ×105. *f*, On the left is a nude mouse showing no evidence of a tumour, 3 weeks after subcutaneous inoculation with 2×10^6 NIH 3T3 cells; on the right is a nude mouse having a large tumour at the site where the mouse was inoculated 3 weeks previously with 5×10^5 cells from a focus transformed by clone 6-2 (human c-Ha-*ras*1 ligated to the LTR of Ha-MuSV). In this experiment none of five mice inoculated with 2×10^6 NIH 3T3 cells developed tumours after 10 weeks, whereas all of five mice inoculated with cells from the lines shown in *b–e* developed palpable tumours within 2 weeks.

5′ end of the gene. Similarly, attempts to induce cellular transformation with the human c-*ras* genes without an LTR have been unsuccessful. In our initial attempts to transform cells with the c-*ras* genes, we have focused on human c-Ha-*ras*1, which preliminary molecular hybridization data have suggested is transcriptionally active in normal human fibroblasts (E.H.C. *et al.*, unpublished data).

We therefore constructed a chimaeric plasmid containing the LTR of Ha-MuSV upstream from human c-Ha-*ras*1 (see Fig. 2). This chimaeric DNA induced focal transformation of NIH 3T3 cells in all of four separate experiments. While the foci induced by Ha-MuSV DNA were usually first detected about seven days after addition of the DNA to the cells, the foci induced by human c-Ha-*ras*1 were generally detected about nine days after addition of the DNA; this short delay has also been observed for rat c-Ha-*ras*1. The results of a typical experiment are shown in Fig. 3. The number of foci induced by a chimaeric plasmid containing an LTR upstream from rat c-Ha-*ras*1 was an order of magnitude lower than that induced by a circularly permuted full-length Ha-MuSV DNA genome. The clone containing human c-Ha-*ras*1 induced only about one-fifth as many foci as did the rat c-Ha-*ras*1 clone. The higher efficiency of transformation by the chimaeric plasmid containing rat c-Ha-*ras*1 may be due to the fact that it carries almost 2 kilobases (kb) more non-*ras* Ha-MuSV sequences upstream from the LTR than were present in the chimaeric plasmid with human c-Ha-*ras*1; deletion of these sequences has previously been associated with a lower efficiency of transformation by Ha-MuSV DNA[5,17]. In further experiments, three independent foci were also obtained in the NIH 3T3 cells when the human c-Ha-*ras*1 fragment (a 2.9)-kb *Sac*I fragment[13]) was ligated without cloning to a 1.4-kb *Sac*I fragment which contained the cloned LTR of the Gardner–Arnstein strain of feline sarcoma virus (FeSV)[18].

In addition to being phenotypically transformed, the cells transformed by human or rat c-Ha-*ras*1 were also tumorigenic for nude mice: palpable tumours developed in less than two weeks in mice inoculated with 5×10^5 transformed cells, whereas no tumours were evident in the 10 weeks following inoculation with 2×10^6 control cells (see Fig. 4*f*).

Human gene transformants contain the LTR and human c-Ha-*ras*1

To determine the presence of the cloned DNA in the transformed NIH 3T3 cells, genomic DNA from transformants was analysed for the presence of LTR and human c-Ha-*ras*1 sequences. Genomic DNAs were digested with restriction endonuclease, and fragments were fractionated by agarose gel electrophoresis, transferred to nitrocellulose paper, and hybridized with ^{32}P-labelled LTR and human c-Ha-*ras*1 probes (Fig. 5). Hybridization to the *ras* probe of *Eco*RI-digested genomic DNA from a transformant yielded two major new fragments of 7.6 and 5.8 kb (Fig. 5*a*, lane 3)—this indicated that the transformant contained additional *ras* sequences. These new fragments were presumably human c-Ha-*ras*1, since at the exposure in the autoradiogram shown in Fig. 5, the probe detected the 25-kb *Eco*RI fragment in human genomic DNA previously identified as containing c-Ha-*ras*1 (ref. 13; Fig. 5*a*, lane 1) and did not reveal endogenous mouse *ras* fragments in the NIH 3T3 cells (Fig. 5*a*, lane 2). The LTR was presumptively identified as being linked to both of these fragments, as hybridization of *Eco*RI-digested DNA from the transformant to the LTR probe detected, in addition to multiple copies of endogenous LTR fragments, two novel fragments which co-migrated with the two fragments identified using the *ras* probe (Fig. 5*b*). The 5.8-kb *Eco*RI fragment is similar in size to the input cloned DNA, suggesting that some of the copies might still be intact circular episomes or integrated in tandem.

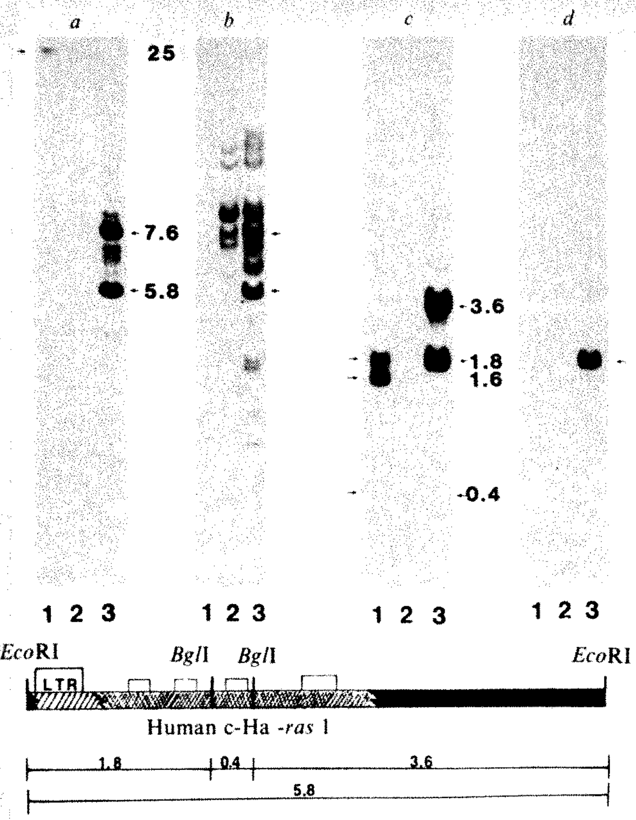


Fig. 5 Southern blot analysis of restriction endonuclease-digested high-molecular weight cellular DNA. Restriction endonuclease-digested genomic DNA was electrophoresed in 0.6% agarose gels, transferred to nitrocellulose filters, and hybridized to c-Ha-*ras* or LTR ^{32}P-labelled probes in 4×SSC, 0.2% of *Escherichia coli*, BSA, PVP, 0.1% SDS and 50 μg μl^{-1} of low-molecular weight salmon testes DNA at 65 °C for 36–48 h (ref. 37). Filters were then washed in 0.1×SSC and 0.1% SDS for 1 h with four changes and in 0.1×SSC for 1 h with two changes at 65 °C. *a*, *b*, All lanes contain *Eco*RI-digested genomic DNAs; *c*, *d*, all lanes contain *Eco*RI/*Bgl*I-digested genomic DNAs. For each panel, lane 1 is human placenta; lane 2, NIH 3T3; lane 3, 6-2 clone (LTR–human c-Ha-*ras*1)-transformed NIH/3T3. The ^{32}P-labelled probes used for hybridization were: human c-Ha-*ras*1, *Sac*I–*Sac*I 2.9-kb insert for *a* and *c*; Ha-MuSV LTR 0.65-kb insert for *b* and *d*. The schematic drawing at the bottom represents the linear molecule resulting from *Eco*RI digestion of the chimaeric plasmid 6-2 or 6-9, described in Fig. 2 legend. Ha-MuSV (non-*ras*) sequences are indicated by shading with diagonal lines, human c-Ha-*ras*1 sequences by cross-hatching, and pBR322 sequences by solid shading. Open rectangles in c-Ha-*ras*1 represent exons.

The linkage of the LTR and human c-Ha-*ras*1 sequences in the DNA of the transformants was confirmed by *Eco*RI+*Bgl*I digestion of the genomic DNAs. This double digestion cleaves the pBR–LTR–human c-Ha-*ras*1 cloned DNA into three fragments: a 1.8-kb fragment containing LTR and some *ras* sequences, a 0.4-kb *ras*-containing fragment, and a 3.6-kb fragment which includes *ras* and pBR sequences (see schematic drawing in lower part of Fig. 5). The 1.8-kb fragment was detected with the LTR and the *ras* probes (Fig. 5*c*, *d*, lane 3), indicating that the LTR and *ras* fragments were linked to each other. The predicted 3.6-kb fragment was also easily detected with the *ras* probe, and the 0.4-kb fragment, not seen at this exposure, was detected in longer autoradiographic exposures. These results indicate that the cloned DNA is intact in the transformed cells. Similar results have also been obtained with genomic DNA from two other foci induced by the cloned DNA (data not shown).

Transformed cells contain high levels of Ha-*ras* p21

Having demonstrated that the LTR and c-*ras* sequences were intact, it was important to determine that the transformed cells contained high levels of Harvey type p21. Immunofluorescent staining of the transformants with a Harvey-specific monoclonal p21 antibody yielded positive results (Fig. 4). While the normal mouse cells were not detectably stained by the antibody, cells transformed by v-Ha-*ras*, rat c-Ha-*ras*1 and human c-Ha-*ras*1 (with FeSV or Ha-MuSV LTR) all gave a similar pattern of intense staining consistent with the membrane fluorescence previously reported for v-Ha-*ras*-transformed cells[19].

The transformants were also analysed by immunoprecipitation for Ha-*ras* p21. When ^{35}S-methionine-labelled extracts of the transformants were reacted with a Harvey p21 monoclonal antibody, high levels of p21 were precipitated (Fig. 6). This was observed both for cells transformed by the chimaeric Ha-MuSV LTR–human c-Ha-*ras*1 plasmid and for a transformant induced by ligation of the FeSV LTR to human c-Ha-*ras*1. The p21 in the cells transformed by the human gene resembled that obtained from cells transformed by rat c-Ha-*ras*1. While the p21 encoded by v-Ha-*ras* consists of two forms, as shown by the double band in the autoradiogram (Fig. 6*b*), the p21 of the cells transformed by either c-Ha-*ras*1 gene contain only a single detectable form of p21. Further details on the biochemical comparison of v-*ras* p21 and c-*ras* p21 will be published elsewhere.

Significance of c-Ha-*ras*-induced transformation

The *ras* multigene family consists of at least two types of p21 genes that were first identified with *ras* probes derived from Harvey and Kirsten murine sarcoma viruses. Two other transforming retroviruses have also transduced c-*ras* sequences: the BALB (B) MuSV, isolated from a BALB mouse[20,21], and the rat MuSV, isolated from rat cells in tissue culture[22,23]. These two viruses have both transduced a Harvey type c-*ras* gene[13,21,23]. While the oncogenic potential of the LTR-activated cloned c-*ras* genes have been demonstrated here only for NIH 3T3 cells, the *ras*-containing retroviruses can transform a wide variety of other cell types from humans and other species, including primary and epithelial cells.

Previous studies have shown that two rodent c-*onc* genes homologous to v-*onc* from a transforming retrovirus can induce oncogenic transformation, when activated by a retroviral LTR leading to their high level expression. The present studies

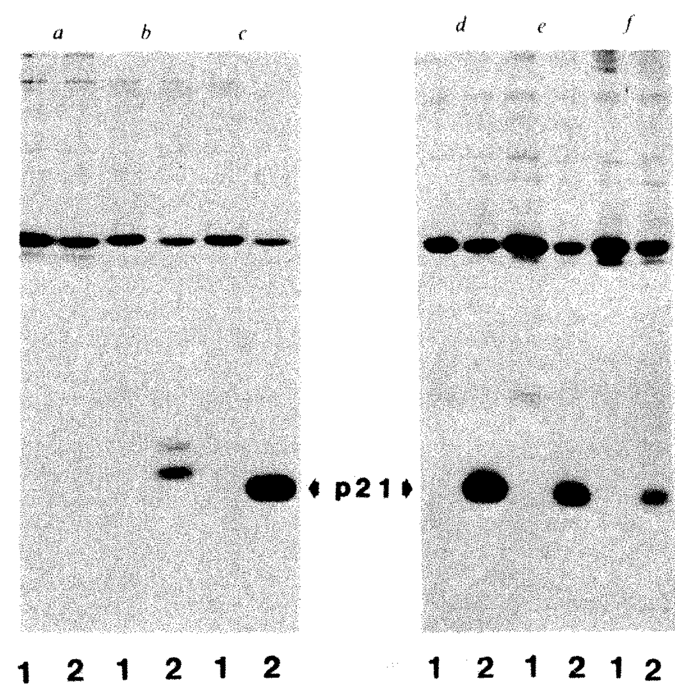


Fig. 6 Gel electrophoresis of immunoprecipitates of p21 in normal c-Ha-*ras*1-transfected NIH/3T3 cells. Cells were labelled with ^{35}S-methionine at 37 °C for 16 h and extracts were prepared and immunoprecipitated as described elsewhere[15,35]. Lane 1, no antibody; lane 2, anti-Harvey p21 monoclonal antibody (Y13-238). *a*, Uninfected NIH/3T3 cells; *b*, Ha-MuSV-transformed cells (Ha 821); *c*, rat c-Ha-*ras*1-transformed cells. *d*–*f*, Human c-Ha-*ras*1-transformed cells: *d*, ligated to FeSV LTR; *e*, clone 6-2; *f*, clone 6-9.

extend this finding to a human c-*onc*, a species from which a transforming retrovirus has not yet been isolated. These results suggest that such transformation may be a general feature of c-Ha-*ras*1 genes. It should also be stressed that these rodent and human genes have been cloned from normal cellular DNA, and they have not by themselves induced focal transformation of NIH 3T3 cells.

Given the tumorigenic properties of these activated normal genes, it seems reasonable to speculate that other cellular oncogenes may possess analogous tumorigenic potential, and that normal transcriptional control mechanisms prevent these genes from being continuously expressed at high levels. These and other examples of LTR-mediated activation of a cellular oncogene[8,9,24–26] suggest that a constitutive change in the transcriptional control of an oncogene could represent an important mechanism in the pathogenesis of some cancers. Such changes in oncogene expression might be induced either by viral or other mechanisms, such as somatic mutation or recombination. In this regard, it has recently been reported that, in contrast to the very low transforming potential of genomic DNA from normal cells, genomic DNA from some human malignancies is able by itself to induce oncogenic transformation of NIH 3T3 cells[27–29]. These transforming sequences from a bladder carcinoma have recently been molecularly cloned and have retained their oncogenicity[30–32]. As reported in the accompanying paper[33], this cloned gene has now been identified as human c-Ha-*ras*1. Similar observations have also been made by others[32,34]. Because the normal c-Ha-*ras*1 gene and the human tumour c-Ha-*ras*1 gene are now available, it should be possible to localize the alterations in the normal gene that are responsible for the dramatic change in biological activity of the gene isolated from the bladder tumour. This analysis may help to elucidate the molecular basis of neoplastic transformation in a human cancer.

We thank Drs Ronald W. Ellis and Thomas Y. Shih for helpful discussions; Mr Jerome Goodwin, Ms Shirley Yu and Ms Marilyn R. Lander; and Ms Barbara Fleurdelys for technical assistance. This work was supported in part by a NCI research support contract to Meloy Laboratories, Springfield, Virginia, USA.

Received 8 April; accepted 30 April 1982.

1. Klein, G. (ed.) *Viral Oncology* (Raven, New York, 1980).
2. Lowy, D. R., Rands, E. & Scolnick, E. M. *J. Virol.* **26**, 291–298 (1978).
3. Andersson, P., Goldfarb, M. P. & Weinberg, R. A. *Cell* **16**, 63–75 (1979).
4. Blair, D. G., McClements, W., Oskarsson, M., Fischinger, P. & Vande Woude, G. F. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3504–3508 (1980).
5. Chang, E. H., Ellis, R. W., Scolnick, E. M. & Lowy, D. R. *Science* **210**, 1249–1251 (1980).
6. Klein, G. (ed.) *Advances in Viral Oncology: Cell Derived Oncogenes* Vol. 1 (Raven, New York, in the press).
7. Oskarsson, M., McClements, W. L., Blair, D. G. & Maizel, J. V. *Science* **297**, 1222–1223 (1980).
8. Blair, D. G. *et al. Science* **212**, 941–943 (1981).
9. DeFeo, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 3328–3332 (1981).
10. Langbeheim, H., Shih, T. Y. & Scolnick, E. M. *Virology* **106**, 292–300 (1980).
11. Gattoni, S., Korschmeier, P., Weinstein, I. B., Escopbedo, J. & Dina, D. *Molec. Cell Biol.* **2**, 42–51 (1982).
12. Chattopadhyay, S. K., Chang, E. H., Ellis, R. W., Scolnick, E. M. & Lowy, D. R. *Nature* **296**, 361–383 (1982).
13. Chang, E. H., Gonda, M. A., Ellis, R. W., Scolnick, E. M. & Lowy, D. R. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
14. Ellis, R. W. *et al. Nature* **292**, 506–511 (1981).
15. Furth, M. E., Davis, L. J., Fleurdelys, B. & Scolnick, E. M. *J. Virol.* (in the press).
16. Ellis, R. W., Defeo, D., Furth, M. E. & Scolnick, E. M. *Cell* (submitted).
17. Chang, E. H. *et al. J. Virol.* **35**, 76–92 (1980).
18. Fedele, L. A., Even, J., Garon, C. F., Donner, L. & Sherr, C. J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4036–4040 (1981).
19. Willingham, M. C., Pastan, I., Shih, T. Y. & Scolnick, E. M. *Cell* **19**, 1005–1014 (1980).
20. Peters, R. L. *et al. J. natn. Cancer Inst.* **53**, 1725–1729 (1974).
21. Andersen, P. *et al. Cell* **26**, 129–134 (1981).
22. Rasheed, S., Gardner, M. B. & Huebner, R. J. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2972–2976 (1978).
23. Young, H. A., Rasheed, S., Sowder, R., Benton, C. V. & Henderson, L. E. *J. Virol.* **38**, 286–293 (1981).
24. Hayward, W. S., Neel, G. B. & Astrin, S. M. *Nature* **290**, 475–480 (1981).
25. Payne, G. S. *et al. Cell* **23**, 311–322 (1981).
26. Neel, B. G., Hayward, W. S., Robinson, H. L., Fang, J. & Astrin, S. M. *Cell* **23**, 323–334 (1981).
27. Shih, C., Shilo, B., Goldfarb, M. P., Dannenberg, A. & Weinberg, R. A. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5714–5718 (1979).
28. Krontiris, T. G. & Cooper, G. M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1181–1184 (1981).
29. Perucho, M. *et al. Cell* **27**, 467–476 (1981).
30. Shih, C. & Weinberg, R. A. *Cell* (in the press).
31. Goldfarb, M., Shimizu, K., Perucho, M. & Wigler, M. *Nature* **296**, 404–409 (1982).
32. Pulciani, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 2845–2849 (1982).
33. Parada, L. F., Tabin, C. J., Shih, C. & Weinberg, R. A. *Nature* **297**, 474–478 (1982).
34. Der, C. J., Krontiris, T. G. & Cooper, G. M. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
35. Shih, T. Y., Weeks, M. O., Young, H. A. & Scolnick, E. M. *Virology* **96**, 64–79 (1979).
36. Graham, F. L. & van der Eb, A. J. *Virology* **52**, 456–461 (1973).
37. Hager, G. L. *et al. J. Virol.* **31**, 795–809 (1979).

LETTERS TO NATURE

Lithium abundance at the formation of the Galaxy

M. Spite & F. Spite

Observatoire de Paris-Meudon, Section d'Astrophysique, 92190 Meudon, France

The observed abundance of lithium in population I stars (young stars)[1] seems to have been stable during the past 5×10^9 yr (ref. 2). This stability has been used previously[1,3] to estimate the abundance of lithium at the time of formation of the Galaxy and at pre-galactic times[1,3]. But because the basis for this estimate was not completely satisfactory, we have made new observations, aiming at a better determination of the lithium abundance at the time of formation of the Galaxy. The newly observed stars are extreme population II dwarfs (very old, very metal-poor stars). Their lithium abundance turned out to be significantly lower than the abundance in population I stars. If attributed to the big bang, this lithium abundance suggests a rather low baryonic density of the Universe, which in turn favours, under some assumptions, an open Universe[3]. Finally, we suggest that the good agreement between the abundance of ^{7}Li and the deuterium ^{2}H abundance[3] supports the standard model of the big bang.

The new Canadian, French, Hawaian (CHF) 3.6-m telescope[4] with its large collecting area, excellent seeing conditions and the sensitivity of its Reticon receiver has made it possible to observe the lithium resonance line[5] in dwarf stars of the extreme population II, often called halo dwarfs (typical representative of the halo of the Galaxy). These stars are very old and their metal deficiency reflects the lack of metals at the time of their formation. They are faint and the lithium line is narrow and faint in such stars, so that its measurement requires high resolution spectra with a high signal-to-noise ratio. Good spectra were obtained for a few of these stars and lines as faint as 3 mÅ can be detected. A comparatively strong lithium line (23–50 mÅ) was observed[5,6] in all except the coolest star. This absence of lithium is easily understood when it is recalled that the convective zone is very deep in cool stars: the lithium is then driven down into very hot layers, where it is completely destroyed by proton fusion[7].

From the measured wavelength of the observed lithium line it seems that the absorbing species is essentially the ^{7}Li isotope.

A classical spectrum analysis using model atmospheres has been carried out to determine the abundance of lithium for each star. The high quality of the spectra means that the error on the equivalent widths is negligible. The random error in determination of the lithium abundance ($n_{\rm Li}$) comes essentially from the lack of precision about the effective temperature ($T_{\rm eff}$) of the stars. For this type of stars it is generally admitted that $\Delta\log T_{\rm eff}=0.01$. If $\log T_{\rm eff}$ is increased by 0.01, then $\log n_{\rm Li}$ will be increased by about 0.2 (Fig. 1).

Clearly, the present sample of Li abundances in population II dwarfs is very small; however, the abundance of lithium is

0028–0836/82/230483—03$01.00

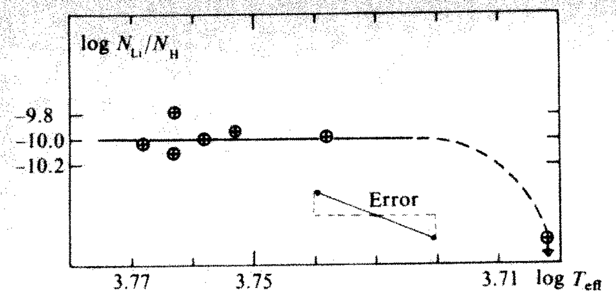


Fig. 1 Lithium abundance plotted against effective temperature (logarithmic scale) for very old metal-poor stars. The lithium abundance is remarkably constant except for the coolest star: here, deep convective movements carry the lithium into very hot layers where it is completely destroyed.

remarkably constant in all the observed halo dwarfs but the coolest one (Fig. 1). This constancy versus the effective temperature T_{eff} of the stars is also a measure of constancy versus the mass of these essentially unevolved stars. If confirmed by more observations, this suggests that lithium is not destroyed by convection in solar type stars of population II, at variance with what is assumed[8] in population I stars. Taking advantage of this constancy, it is easy to show that the lithium abundance (relative to hydrogen) in the atmospheres of old halo dwarfs is $N_{Li}/N_H = 10^{-10}$—about 10 times less than the population I abundance[1].

This lithium is not expected to come from the inside of the star: nucleosynthesis of lithium has been predicted only for stars in a much more advanced stage of evolution[9]. Nor is the ^{7}Li expected to come from outside: most of these are stars whose velocity, relative to population I interstellar clouds, is high, so that the accretion process is inefficient[10,11].

The lithium observed in the halo dwarfs has its origin in the lithium existing in the material which formed these stars. The abundance of lithium found in the atmosphere of the star should not be very different from the abundance in this material. Admittedly, several causes of depletion of lithium may exist: some lithium may be destroyed during the contraction of the protostellar cloud[7,12] (pre-main sequence phase); microscopic and/or turbulent diffusion (acting during the very long life in the main sequence of these very old stars) could deplete the lithium[13]; the stellar mass-loss acting over a long time could also cause depletion (H. Reeves, personal communication). All these effects are not yet well known for halo stars; moreover, at first sight it could be conjectured that they are mass (and/or temperature) dependent; this was confirmed by other observations, and the constancy of the lithium abundance suggests that these processes are not very efficient. Keeping in mind the causes of uncertainty, let us assume that $N_{Li} = 10^{-10}\ N_H$ is the abundance of the material which formed the halo dwarfs. This abundance cannot be significantly different from the abundance of the pre-galactic material: admittedly, the atmosphere of the halo dwarfs includes some material processed in supernovae; this material, when astrated, was on the one hand deprived of its lithium, and on the other hand strongly enriched in metals. As the halo dwarfs are metal poor, only a very small fraction of the astrated material can have been mixed in the pre-galactic material; the lithium abundance seems constant within rather wide limits of metal abundance (1/20–1/200 of the solar metal abundance). It seems therefore that we can take the lithium abundance of the atmospheres of the halo dwarfs as the abundance of the pre-galactic material.

Evaluated by mass, this pre-galactic lithium abundance is $X_7 = 5.2 \times 10^{-10}$ (assuming a helium abundance $Y = 0.25$ by mass). If attributed to the big bang, this abundance leads[14], according to the standard theory, to a rather low baryonic density of the Universe (Fig. 2) lower than the density previously derived by Boesgaard[1] and by Austin and King[3]. The density found remains, however, in agreement with that derived from the abundance of the deuterium[3], within the estimated errors. For an easier comparison with the work of Austin and King, the Hubble constant retained was $H_0 = 55$ km s^{-1} Mpc^{-1}.

The error of the primordial lithium abundance determination may be estimated as follows.

(1) The classical abundance analysis, using models of atmospheres, with a few simplifying assumptions (such as local thermodynamical equilibrium), may induce an error as high as a factor of 2 (about $10^{0.3}$ or 0.3 dex). However, the systematic part of this error is about the same for populations I and II, so that it cancels out, at least partly, when comparing relative abundance in populations I and II. When using several stars, the random part of the error decreases accordingly, even when adopting the parameters T_{eff} and g as they appear in the literature (without homogenization).

(2) There is some uncertainty about the possible depletion of lithium, owing to the possible proto-stellar destruction and/or depletion due to diffusion, and/or lithium depletion due to stellar wind. This uncertainty is difficult to estimate. We can adopt, as an upper limit, a factor of ~3 (0.5 dex), because Duncan[8] and Bodenheimer[12] predict only a factor of 2 for proto-stellar destruction for solar type stars.

Thus taking account of the cumulated errors, the final uncertainty about the difference between the lithium abundance in populations I and II is probably less than a factor of 6 (0.8 dex); so that we have to accept that the abundance of lithium is probably lower in population II than in population I.

It seems that the evolution of the Galaxy has not significantly decreased the abundance of the primordial lithium[2]; as lithium is destroyed inside stars and stellar envelopes, a lithium-producing process must be found. Spallation is only partially responsible for this production[15]; a promising possibility is production in novae[16].

Even considering the uncertainties involved in the interpretation of the lithium abundance in halo dwarfs, our work provides a clear demonstration that the Universe cannot be closed by

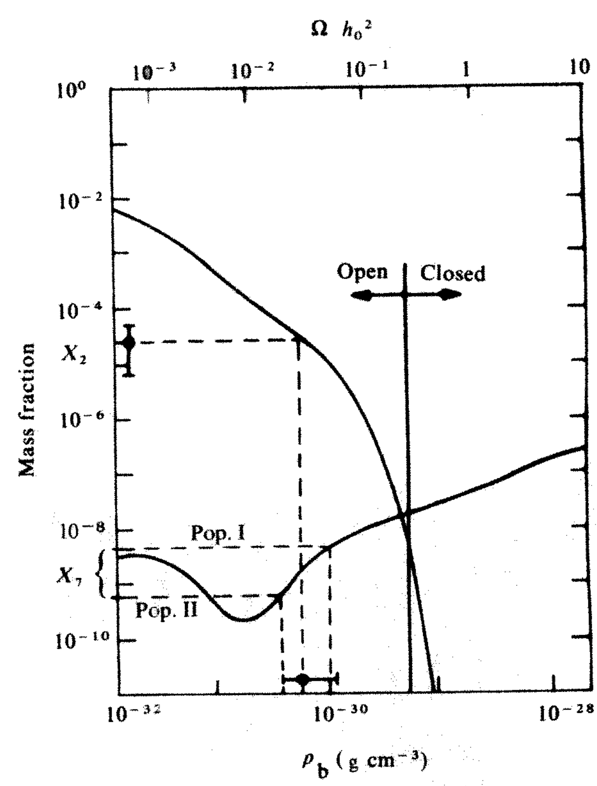


Fig. 2 The new lithium abundance X_7 found in this work leads to a rather low baryonic density, in agreement (within determination errors) with that derived from deuterium abundance. The values of the Hubble constant is the same as in ref. 3. If the cosmological constant Λ is assumed to be zero, this leads, more clearly than in previous work, to the conclusion that the baryonic density is not large enough to close the Universe.

nucleons, when assuming the validity of the standard model[14] of the big bang and assuming a zero value for the cosmological constant ($\Lambda = 0$). This result remains unchanged when compared with the model of Yang *et al.*[17], which has three pairs of neutrinos ($\Omega_N h_0^2 = 2 \times 10^{-2}$). More observations, a better theory of protostellar and stellar convection, a better theory of diffusion and a better theory of stellar wind are needed to obtain a more accurate value of the primordial lithium abundance. However, even within the present uncertainty, some agreement exists between the baryonic density predicted from lithium and deuterium abundances, and when confirmed, this agreement will support the validity of the standard model of the big bang.

We thank the CFHT staff and especially Drs B. Campbell, G. Lelièvre, J-P. Maillard and Mr B. Grundseth for help with the observations. The data were processed through the VAX computer of the Service Informatique of the Paris–Meudon Observatory: we thank especially Mr Tran Van Tue for help with tape decoding.

Received 14 December 1981; accepted 26 March 1982.

1. Boesgaard, A. M. *Publ. astr. Soc. Pacif.* **88,** 353–366 (1976).
2. Audouze, J. & Tinsley, B. M. *Astrophys. J.* **192,** 487–500 (1974).
3. Austin, S. M. & King, C. H. *Nature* **269,** 782 (1977).
4. Cayrel, R. *J. Opt.* **11,** 5–6 (1980).
5. Spite, M. & Spite, F. *C.r. hebd. Séanc. Acad. Sci., Paris* **293,** 299–301 (1981).
6. Spite, M. & Spite, F. *Astr. Astrophys.* (submitted).
7. Boesgaard, A. M. in *Highlights of Astronomy* Vol. 4, Part II (ed. Müller, E. A.) 209–216 (Reidel, Dordrecht, 1977).
8. Duncan, D. K. *Astrophys. J.* **248,** 651–669 (1981).
9. Sneden, Ch., Lambert, D. L., Tomkin, J. & Peterson, R. C. *Astrophys. J.* **222,** 585–594 (1978).
10. McCrea, W. H. *Mon. Not. R. astr. Soc.* **113,** 162–179 (1953).
11. Talbot, R. J. Jr & Newman, M. J. *Astrophys. J. Suppl. Ser.* **34,** 295–308 (1977).
12. Bodenheimer, P. *Astrophys. J.* **142,** 451–461 (1965).
13. Vauclair, S., Vauclair, G., Schatzman, E. & Michaud, G. *Astrophys. J.* **223,** 567–582 (1978).
14. Wagoner, R. V. *Astrophys. J.* **179,** 343–360 (1973).
15. Meneguzzi, M., Audouze, J. & Reeves H. *Astr. Astrophys.* **15,** 337–359 (1971).
16. Starrfield, S., Truran, J. W., Sparks, W. M. & Arnould, M. *Astrophys. J.* **222,** 600–604 (1978).
17. Yang, J., Schramm, D. N., Steigmann, G. & Rood, R. T. *Astrophys. J.* **227,** 697–704 (1979).

Five-minute oscillations as a subsurface probe of sunspot structure

John H. Thomas*, Lawrence E. Cram† & Alan H. Nye‡

* University of Rochester, Rochester, New York 14627, USA
† Sacramento Peak Observatory, Sunspot, New Mexico 88349, USA
‡ Rochester Institute of Technology, Rochester, New York 14623, USA

Observations are reported here which show that the 5-min oscillations in a sunspot umbra actually split into several individual modes of different period. We interpret these modes of oscillation as the response of the sunspot to forcing by the 5-min p-modes in the surrounding quiet atmosphere. Also, we show how detailed observations of the multiple 5-min modes in a sunspot may be used as a probe of the structure of a sunspot beneath the solar surface.

Oscillatory velocity fields in sunspots may be divided into two categories. The first category consists of the umbral oscillations (period 145–190 s) and the running penumbral waves (period 180–250 s). Observations suggest that these waves are resonant modes (see refs 1, 2), and theoretical work[3–7] has identified these waves with resonant magneto-atmospheric wave modes of the sunspot atmosphere itself. These waves, which are presumably excited by subsurface oscillatory convection, offer indirect evidence of the existence of overstable convective modes in the subphotosphere.

The other category of oscillations in sunspots consists of oscillations with periods around 5 min, the dominant period of oscillations in the quiet photosphere. Recent observations[8–10] have overcome the problems of stray light from the surrounding photosphere and the lack of an absolute wavelength reference and have confirmed the existence of 5-min oscillations in sunspot umbras, at power levels well below that in the quiet photosphere. As there are no resonant modes of sunspot oscillations in the 5-min range, we have suggested[11] that these 5-min oscillations represent the passive response of the sunspot to forcing by the 5-min p-mode oscillations in the surrounding gas. It is now well established that the 5-min oscillations in the quiet photosphere actually consist of several modes of oscillation, in the form of acoustic waves trapped in the upper convection zone (p-modes)[12]. These p-modes include a wide range of horizontal wavelengths with a moderate variation of frequency. It is expected that the geometry of the sunspot and the nature of the coupling between the acoustic waves and the sunspot flux tube will act as a filter, with only certain p-modes forcing the sunspot with observable amplitude.

The present observations were made as part of a large programme to study dynamical phenomena in sunspots, using the vacuum tower telescope and Echelle spectrograph at Sacramento Peak Observatory. Here we show only the results of two high-quality time series of velocity oscillations in the umbral photosphere measured in the spectral line Fe I λ6,302.5. The velocity was measured from photographic spectra using the Doppler shift of the magnetic V profile of the line, thus assuring that the origin of the signal is from within the sunspot and not due to scattered light. Also, a nearby telluric line of

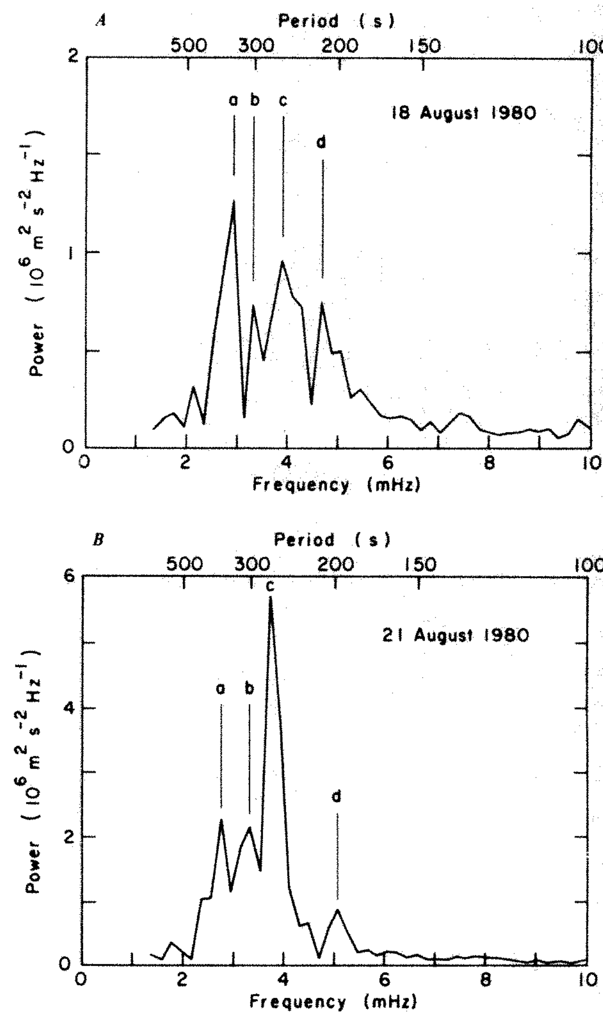


Fig. 1 The mean power spectrum of line-of-sight motions in the umbral photosphere for two runs: 18 August 1980 (*A*) and 21 August 1980 (*B*). The peaks have the following periods. *A*: a, 353 ± 6 s; b, 301 ± 4 s; c, 263 ± 3 s; d, 213 ± 5 s; *B*: a, 366 ± 6 s; b, 301 ± 4 s; c, 269 ± 4 s; d, 197 ± 4 s. Peaks a, b and c are interpreted as representing the response of the sunspot to individual p-mode oscillations in the surrounding atmosphere.

0028–0836/82/230485—03$01.00

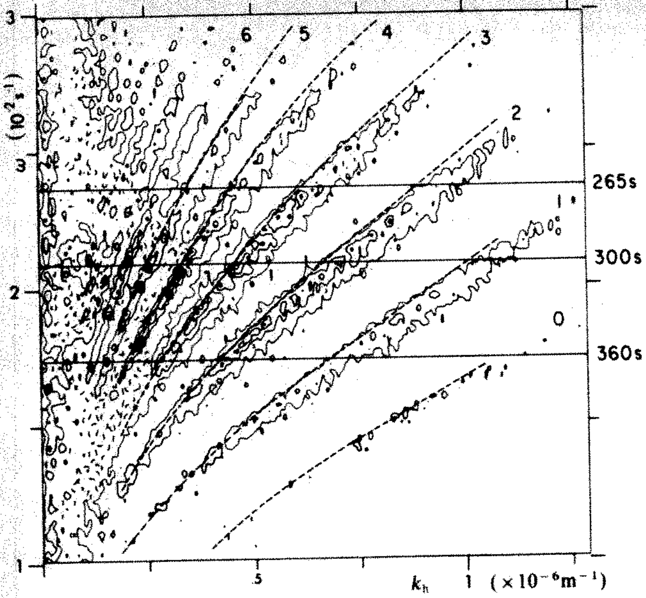


Fig. 2 Highly refined diagnostic diagram for the solar p-modes, due to Deubner *et al.*[14], showing contours of power on the frequency–horizontal wavenumber plane. Lines corresponding to the nominal periods of umbral oscillation near 5 min (360 s, 300 s and 265 s) are drawn to determine their points of intersection with p-modes of different radial mode number n (also shown).

O_2 was used as a wavelength reference, thus avoiding the dangers inherent in using a photospheric line as a velocity reference.

Figure 1 shows the mean power spectrum of the velocity field for the portion of the umbra covered by the entrance slit, for each of two runs (18 and 21 August 1980). Each power spectrum has three conspicuous peaks clustered around a 5-min period. Given the frequency resolution of our observations, the periods of these three peaks for both runs are consistent with nominal values of 265 s, 300 s and 360 s. Another conspicuous peak near 200 s in both runs is more likely to be associated with a resonant mode of the sunspot itself. The greater power in the peaks on 21 August, when the sunspot crossed the central meridian of the Sun, indicates that the velocity is predominantly vertical in the photosphere. The total power at periods near 5 min implies an r.m.s. vertical velocity of $\sim 50\ m\ s^{-1}$ in the umbra, at least a factor of five smaller than for the 5-min oscillation in the quiet photosphere.

Because of the limited temporal and spatial extent of the data (70-min time series with a 20-s sampling interval, over about 13×10^6 m in the umbra) there is statistical noise in the power spectrum. The noise level may be estimated in three ways: from the high-frequency limit of the power spectrum, from the statistics of the distribution of individual power spectra that are used to form the mean power spectrum for the entire umbra, and from the estimated number of degrees of freedom (~20) using Edmond's tables[13]. All three of these methods indicate that the 50% confidence level for each peak corresponds to roughly 25% of the peak amplitude. Any peak that is four times the noise level is marginally significant (50% confidence level), with greater confidence for higher peaks. Thus, in Fig. 1, the peaks near 360 s, 300 s, 265 s, and 200 s are all significant at well above the 50% confidence level. Note that the noise level is about the same on the two days while the signal increases in power on 21 August as the line-of-sight velocity becomes more nearly vertical.

We interpret the three periods of oscillation around 5 min (265 s, 300 s, and 360 s) as the response of the sunspot to three distinct modes of 5-min oscillation in the surrounding photosphere. We pursue this by plotting these three periods, in Fig. 2, on a highly refined k–ω (wavenumber–frequency) diagram for the photospheric oscillations (due to Deubner *et al.*[14]). Note that the three umbral periods cross all of the p-mode ridges in the k–ω diagram. We may identify the three umbral periods as the response of the sunspot to three consecutive p-modes, but there is ambiguity in choosing the correct set of p-modes. Each set of three mode crossings corresponds to a set of three horizontal wavenumbers and three corresponding horizontal wavelengths of p-mode oscillation. One might expect some relation between the diameter of the sunspot umbra and the wavelength of peak response to the surrounding p-mode excitation. The observed sunspot umbra is irregular in shape and has an area of $\sim 170 \times 10^{12}\ m^2$, giving an effective diameter (diameter of a circle of the same area) of 14.7×10^6 m. (This estimate is subject to the uncertainty in locating the umbral–penumbral boundary by photographic methods). This surface umbral diameter falls within the range of possible wavelengths. However, because the diameter of the sunspot should decrease with depth (due to increasing pressure in the surrounding convection zone), the surface diameter is probably not an appropriate value for comparison.

We suggest that observations of the multi-mode structure of 5-min oscillations in sunspots may be used as a probe of sunspot structure below the solar surface. The various p-modes have eigenfunctions with different depth dependence, with higher modes extending deeper into the convection zone, so that each mode 'probes' a different range of depths. This fact has already enabled the p-modes to be used as a probe of radial variations in the solar rotation rate in the convection zone[14–16]. Here we suggest a similar technique to probe the structure of a sunspot below the surface by measuring the response of the sunspot to the various p-modes. This method calls for accurate observations of the period and spatial structure of the various 5-min modes in the sunspot as well as theoretical work on the interaction between the p-modes and the sunspot flux tube.

An illustration of such a depth probe can be given on the basis of the present observations and a simple theoretical argument. Space–time plots of the umbral velocity field show a characteristic 'herringbone' structure with phase varying across

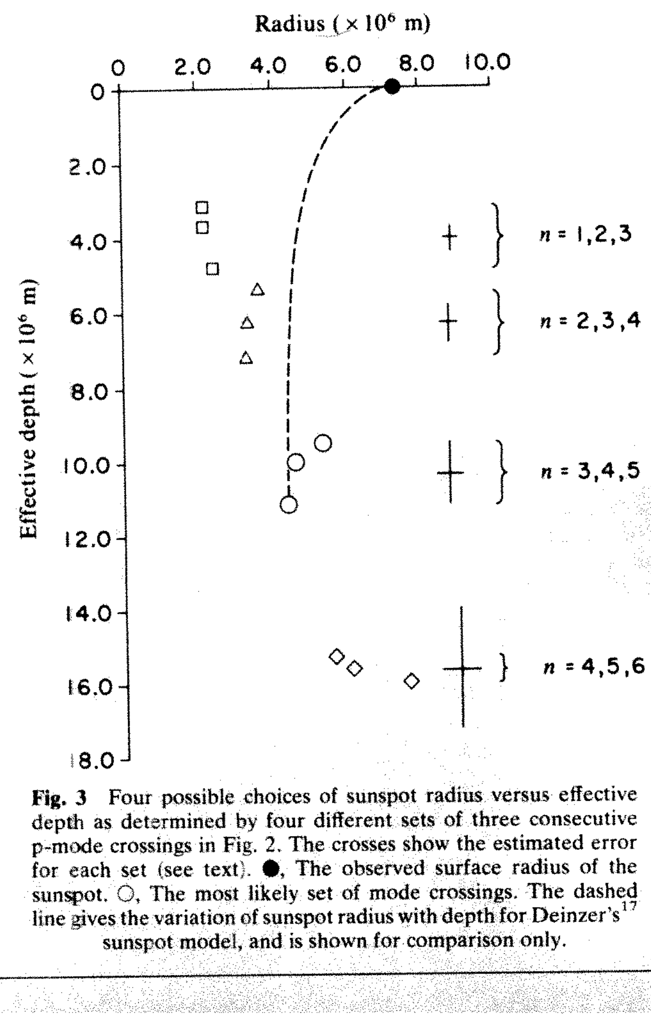


Fig. 3 Four possible choices of sunspot radius versus effective depth as determined by four different sets of three consecutive p-mode crossings in Fig. 2. The crosses show the estimated error for each set (see text). ●, The observed surface radius of the sunspot. ○, The most likely set of mode crossings. The dashed line gives the variation of sunspot radius with depth for Deinzer's[17] sunspot model, and is shown for comparison only.

the umbra. At any instant, diametrically opposite points at the edge of the umbra tend to be 180° out of phase and the centre of the umbra appears to be a node in the velocity field. This suggests that the peak response is occuring at a wavelength equal to roughly twice the umbral diameter. We will assume that the peak response is at a wavelength equal to exactly twice the local diameter and proceed. As discussed above, the various sets of three consecutive mode crossings in Fig. 2 predict alternative sets of three horizontal wavelengths and hence sets of three values of sunspot diameter. Each diameter corresponds to a different p-mode (different value of n) which probes a different effective depth. Values of the effective depth for the p-mode eigenfunctions have been calculated by Ulrich *et al.*[16]. The effective depth varies with horizontal wavenumber as well as radial mode number. For any set of mode crossings in Fig. 2 we can assign an effective depth to each of the three values of sunspot diameter and thus obtain a plot of sunspot radius versus depth, assuming a circularly symmetric sunspot flux tube.

The results of these calculations are combined in Fig. 3, where sunspot radius is plotted as a function of depth for each of the possible sets of mode crossings. The limited resolution in frequency (due to finite time series) and the spread of power in the ridges on the k–ω diagram give the uncertainty in calculated sunspot radius, and the uncertainty in k and ω gives the uncertainty in effective depth. For any given set of points in Fig. 3 the uncertainties mask the depth variation of sunspot radius, making each set of three points equivalent to a single piece of information about the subsurface sunspot geometry. The individual sets of three points in Fig. 3, corresponding to sets of three consecutive mode crossings, are distinct from each other, allowing a selection of the most likely set based on theoretical arguments about the expected depth variation of sunspot radius. For example, the theoretical model of Deinzer[17] is plotted in Fig. 3 for comparison and passes through the set of points corresponding to p-modes $n = 3$, 4 and 5. We use this only as an illustration of the method and not as a specific prediction for this sunspot. However, for the range of wavenumbers under consideration, modes $n = 3$, 4 and 5 have the most power in their ridges; they might be chosen on this basis as well.

J.H.T. was supported by NSF grant ATM-8021305 and NASA grant NSG-7562. A.H.N. was a USAF-SRFP Research Associate at Sacramento Peak Observatory.

Received 7 January; accepted 22 April 1982.

1. Moore, R. L. *Space Sci. Rev.* **28,** 387–421 (1981).
2. Moore, R. L. in *The Physics of Sunspots* (eds Cram, L. E. & Thomas, J. H.) 259–311 (Sacramento Peak Observatory, 1981).
3. Nye, A. H. & Thomas, J. H. *Solar Phys.* **38,** 399–413 (1974).
4. Uchida, Y. & Sakurai, T. *Pub. astr. Soc. Jap.* **27,** 259–274 (1975).
5. Nye, A. H. & Thomas, J. H. *Astrophys. J.* **204,** 582–588 (1976).
6. Scheuer, M. A. & Thomas, J. H. *Solar Phys.* **71,** 21–38 (1981).
7. Thomas, J. H. & Scheuer, M. A. *Solar Phys.* (in the press).
8. Soltau, D., Schröter, E. H. & Wöhl, H. *Astr. Astrophys.* **50,** 367–370 (1976).
9. Livingston, W. & Mahaffey, C. in *The Physics of Sunspots* (eds Cram, L. E. & Thomas, J. H.) 312 (Sacramento Peak Observatory, 1981).
10. Nye, A. H., Cram, L. E., Thomas, J. H. & Beckers, J. M. in *The Physics of Sunspots* (eds Cram, L. E. & Thomas, J. H.) 313–317 (Sacramento Peak Observatory, 1981).
11. Thomas, J. H. in *The Physics of Sunspots* (eds Cram, L. E. and Thomas, J. H.) 345–358 (Sacramento Peak Observatory, 1981).
12. Deubner, F.-L. in *The Sun as a Star* (ed. Jordan, S.) 65–84 (US Government Printing Office, Washington DC, 1981).
13. Edmonds, F. N. *Astrophys. J.* **144,** 733–753 (1966).
14. Deubner, F.-L., Ulrich, R. K. & Rhodes, E. J. Jr *Astr. Astrophys.* **72,** 177–185 (1979).
15. Rhodes, E. J. Jr, Deubner, F.-L. & Ulrich, R. K. *Astrophys. J.* **227,** 629–637 (1979).
16. Ulrich, R. K., Rhodes, E. J. Jr & Deubner, F.-L. *Astrophys. J.* **227,** 638–644 (1979).
17. Deinzer, W. *Astrophys. J.* **141,** 548–563 (1965).

Direct gravimetric detection of magma movements at Mount Etna

T. J. O. Sanderson

Geophysics Section, Imperial College of Science and Technology, London SW7 2BP, UK

Mount Etna is Europe's most active volcano, erupting some 2×10^{10} kg of new material each year[1]. Its eruptions generally have two distinct forms: continuous explosive activity near the summit with low rates of lava effusion, or massive, and often quieter, lava effusion from its flanks along regional tectonic fissures. The question of where magma is stored before eruption is not yet resolved. Here I present results from high-precision gravity surveys covering two recent eruptions. Gravity changes after the major flank eruption of March 1981 indicate draining of magma from a dyke 14 km long, at an elevation of 0–1 km and striking in a direction close to that of the eruptive fissure. Changes from the explosive sub-terminal eruptions of September 1980 do not indicate draining from any near-surface reservoir. It is proposed that short-term magma storage can occur close to the surface before a flank eruption, but that summit eruptive activity is consistent with a source at greater depth.

Recent research on Mount Etna has shown that it does not conform to the classic inflation–deflation model[2] which can be applied successfully to many other volcanoes, notably Kilauea[3,4]. Precise levelling and dry tilt studies show no gross systematic deformation, but only local anomalies[5], and horizontal deformation observed in the summit region is inconsistent with a simple sub-surface reservoir[6]. Furthermore, petrological studies on historical lava flows indicate little of the chemical differentiation which would be expected if substantial high-level magma storage were involved[7,8], and seismic investigations indicate a large volume of partial melt at a depth of 16–24 km beneath the volcano[9]. It has been concluded that magma rises directly from great depth to the surface without long-term storage in a magma chamber[10–12]. It is accepted, however, that high-level storage must take place in the short term, because all lavas show extensive crystallization of plagioclase, indicating residence at moderate to low pressures[12], and some flank effusions are only mildly explosive, indicating degassing before eruption. The lavas of 1910, 1928, 1950–51 and 1971 also show a limited degree of magmatic evolution due to differentiation[13–14]. This implies high-level storage on a time scale sufficient to give rise to crystallization and degassing, but insufficient to result in strong differentiation. To address the problem of where magma resides before eruption, a high precision gravity network was established on Mount Etna in September–October 1979 (ref. 15) with the aim of detecting small secular changes in the force of gravity as a result of intrusion or draining of matter.

Ninety-seven gravity stations have been set up, of which 48 are used in the surveys under discussion (Fig. 1*a*). They are sited on stable massive lava outcrops or permanent architecture. Station relocation is good to ±1 cm, and vertical reference to ±1 mm. Instrument orientation is constant at each station. Measurements are carried out with LaCoste and Romberg gravity meter G90, fitted with electrical readout. Tidal corrections are made using a harmonic tidal programme based on Cartwright and Tayler[16,17] and using gravimetric factors and anomalous phases determined for Mount Etna by Mohamed[18]. Tares and drift are removed using a least-squares optimization programme. Gravity values are referred daily to a base station at the Astrophysical Observatory at Serra la Nave and the network as a whole is referred to a datum based on this and calibration stations numbers 59, 61 and 64 of the Italian Geodetic Commission. The typical standard deviation of a single gravity tie with meter G90 is found to be 9 μGal, and the errors are normally distributed. Each station tie is performed if possible eight times on independent occasions, and it is estimated that calibration variations, datum inaccuracies and water-table effects contribute a further nominal error of 5 μGal. The resulting error for a gravity change is typically 8–15 μGal, depending on the number and quality of the ties made. In the interval September 1979–August 1980 gravity changes in the range −36 to +48 μGal were observed. A χ^2 test gives a probability of <1% that they are due to chance.

0028-0836/82/230487—04$01.00

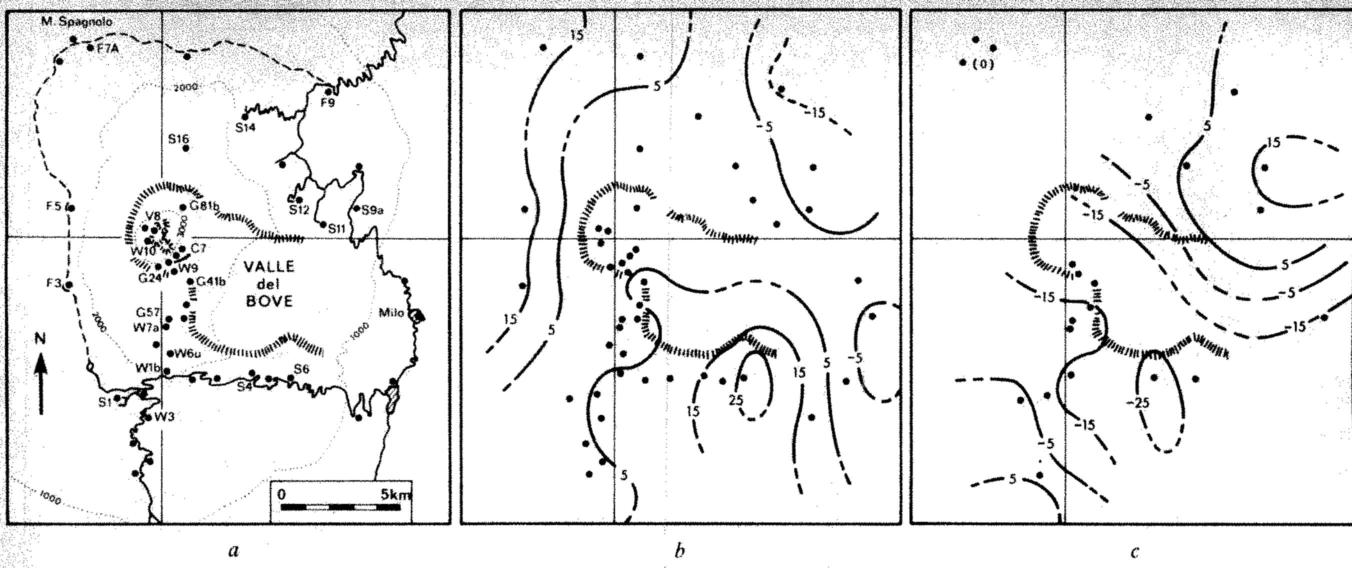


Fig. 1 *a*, Map of Mount Etna: ●, gravity stations; hatching, caldera rims. Topographic contours are in intervals of 1,000 m. *b*, Smoothed contours to show broad features of gravity changes between August and September 1980. Contours are plotted by computer, using a routine in which the gravity change value at any point is calculated as a weighted average of all measurements within a circular window around it. Measurements are weighted inversely to variance and the square of distance from any station, and are smoothed using a filter of cutoff wavelength 0.750 km. A window of radius 5 km is used. Contours are at intervals of 10 μGal, and are dashed where information is poor. Details of the contours should not be taken seriously. *c*, Smoothed contours to show broad features of gravity changes September 1980–March 1981.

A change in the measured force of gravity can be due to two effects. First, an uplift of terrain leads to a reduction in gravity according to the inverse square law. This reduction is ~ 0.22 mGal m^{-1} and can be used for the detection of vertical elevation changes[19]. Second, it can be due to sub-surface density changes or movements of matter. To distinguish between these effects 12 of the gravity stations are common to precise levelling stations of Murray and Guest[5]. They were measured simultaneously with the levelling traverse in September 1979 and July 1980. During this period gravity changes of between −36 and 13 μGal were observed at stations for which the elevation changes were respectively −0.6 cm and 0.4 cm (J. B. Murray, personal communication). If the gravity changes were due to elevation changes the corresponding elevation changes would be +16.4 cm and −5.9 cm. The observed changes are not only too small by an order of magnitude, but they are of opposite sense. This result, and the fact that some 35 dry tilt stations covering the volcano show no pattern of inflation or deflation[5], is a justification for interpreting the gravity changes as due to sub-surface accumulation of magma without corresponding ground deformation.

Such accumulation, whether the magma reservoir is in the form of a dyke system or a cylindrical conduit, can occur in several ways. If the reservoir is connected to an open surface, degassing to the atmosphere occurs, leaving the magma more dense. If the system is closed, it can accommodate the introduction of matter through bubble compression[20] or through bulk compression[21]. Whether the system is open or closed, magma may be introduced at the expense of trapped gases, which are either compressed, absorbed or simply expelled. Mount Etna can be considered, on the basis of tectonic considerations and observation of surface fissure trends, as a system of vertical radial dykes extending along major regional faults[22]. It is proposed here that magma is rising passively into vertical fissures in which the gas pressure and magma level are variable. The magma is not intruding actively through consolidated rock and causing great elastic strain; it is effectively filling a void. Eruption can then occur as a result of tectonically controlled block movements and swelling of magma as pressure is released[22]. If limited surface deformation does occur, as was observed for 2 yr before the 1978 eruption[5,12], this is due to stresses associated with the changing level of magma in the fissure, but the strain involved is not a reflection of swelling equal in volume to the magma newly introduced. A similar, though less extreme, discrepancy has been observed between surface deformation and apparent subsurface volume change on Kilauea[23].

The observed gravity changes on Mount Etna in the period 1979–80 are readily interpretable in terms either of known summit morphology changes or of density changes in historical fissure zones[24]. Dykes involved in the modelling are typically some 500 m below the surface, of vertical extent $\sim$1,000 m and of width 1–5 m. However, the parameters are poorly constrained, as the amount of matter gained or lost is not known independently.

In the case of an observed eruption accurate estimates can be made of the amount of material lost from the volcano. Known gravity changes then enable tight constraints to be placed on regions from which it could have come. Consider the

Table 1 Measured gravity changes (μGal) for intervals August–September 1980 and September 1980–March 1981, with associated errors

Location	Station	Gravity change August–September 1980	Error	Gravity change September 1980–March 1981	Error
Rifugio Sapienza	W1b	7	7	−19	7
	W6u	−1	8	—	—
	G57*	−6	8	−11	8
Belvedere	G41b*	10	8	−22	8
Torre del Filosofo	W9*	8	7	−20	8
Vulcarolo	C7*	0	16	−13 (−25)	17
	G24*	3	8	—	—
	W10*	−9	8	—	—
Summit car park	V8	−9	9	—	—
	G81b*	−7	10	—	—
Mount Pizzillo	S16	−3	10	—	—
Rifugio Puchoz	S14	6	8	−3	8
Rifugio Citelli	S12	−8	12	—	—
	S11	2	8	—	—
	S9a	−11	12	13	12
	S6	35	11	−9	13
	S4	19	9	−33	8
	W3	14	10	—	—
Serra la Nave	S1	−3	7	4	8
	F3*	29	10	—	—
	F5*	26	8	—	—
Mount Spagnolo	F7a*	17	13	−4 (5)	8
da Filippo	F9*	−22	13	8	11

Only those stations measured with accuracy 12 μGal or better are included. The two results in parentheses are corrected for elevation (see text).
* Elevation control exists.

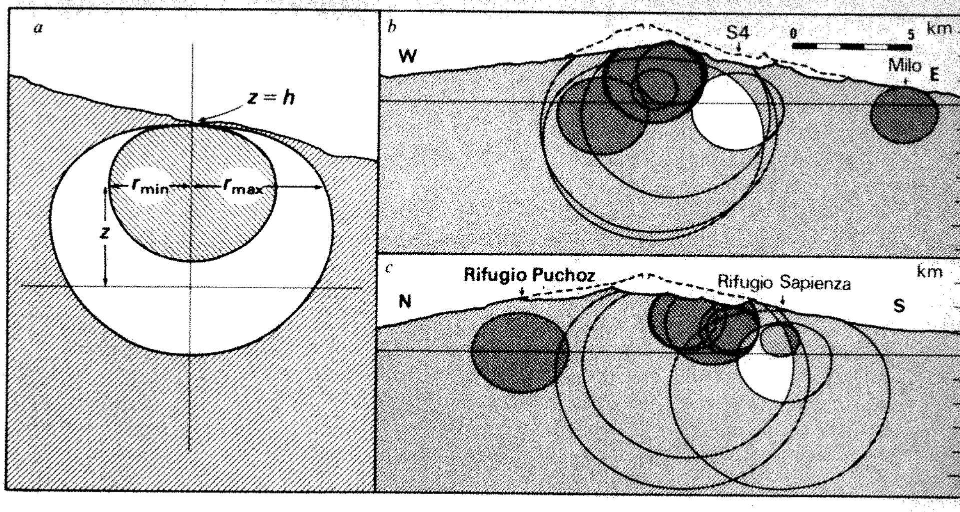


Fig. 2 *a*, Forbidden regions for point mass magma source. Hatching from top left to bottom right indicates a region too close to a station; hatching from bottom left to top right indicates a region too far from one. Sections are shown: *b*, W–E, 4 km south of the summit and *c*, N–S, 2 km east of the summit. The summit profile is shown as a dashed line. The blank region is permissible.

eruption of March 1981. The eruption was only mildly explosive, and followed a series of smaller but highly explosive eruptions during the winter from the North-East Crater[25]. On 17 March 1981, a fissure opened in the north-west flank of the volcano at elevation 2,250 m and gave rise to strong lava fountaining and massive lava effusion. During the next 6 days the fissure system lengthened to 7.5 km in a north-northwest direction and gave rise to an areal coverage of 4.72 km^2 of lava, which reached the River Alcantara at 800 m elevation and threatened the town of Randazzo (Fig. 3*d*). The total volume of lava is estimated at 18.6×10^6 m^3, a mass of about 4.7×10^{10} kg (estimate by J. B. Murray, from mapping by J. E. Guest, C. R. J. Kilburn, R. M. C. Lopes, J. B. Murray, H. Pinkerton and S. Scott).

A gravity survey was made immediately after the eruption and compared with readings taken in September 1980. The results are corrected for the attraction of new lavas lying on the surface. All those results with error bars better than 12 μGal are tabulated in Table 1 with associated errors, and are plotted as smoothed contours in Fig. 1*b*. They show a substantial decrease at all stations high on the southern flank. Towards the region of Rifugio Citelli observed changes were positive or insignificant and at a station 0.5 km from the rift itself there was no significant change. Snow cover prevented measurement of stations in the northern upper reaches and in the Valle del Bove. Snow cover also prevented elevation change control. However, precise levelling surveys have been carried out in August 1981. Two gravity stations showed unusually high elevation changes: station C7, close to the central craters, showed −4 cm (J. B. Murray, personal communication), and station F7a, 0.5 km from the rift, showed +3 cm (M. Grimaldi, personal communication). None of the other stations discussed here showed changes of more than 1.5 cm. As these measurements were made over the interval July/August 1980–August 1981 they are not strictly comparable with gravity changes measured between September 1980 and March 1981, and have not therefore been used to apply an elevation correction to the gravity results. The effect of applying such a correction to results at stations C7 and F7a is included in Table 1, and does not materially alter the picture.

Let us assume that the changes are due to the eruption, and begin with the hypothesis that the erupted magma came from one localized region within the volcano. The magma is equivalent to a sphere of radius 160 m and can be treated as a point mass for depths of the order of 1 km. Now, at station G41b, for example, the measured gravity change was −22 μGal ± 8. At the 95% confidence level the gravity change therefore lies between −38 and −6 μGal. This range of possible values defines 'forbidden' regions from which the magma could not have come. If a station elevation is h above sea level and the total mass erupted is M, then magma loss is permitted only from a region of radius r from a vertical axis beneath the station (Fig. 2*d*), where r is a function of elevation z, and satisfies the inequality:

$$\left(GM\frac{(h-z)}{(\Delta g-2e)}\right)^{2/3}-(h-z)^2<r^2<\left(GM\frac{(h-z)}{(\Delta g+2e)}\right)^{2/3}-(h-z)^2$$

Here G is the universal gravitational constant, Δg is the measured gravity change and e is the error associated with it. M and Δg are negative. The inequality defines spheroidal lobes of radius r_{min} and r_{max} beneath the station. Magma loss is permissible from any point between the two lobes. A three-dimensional envelope of lobes for all stations indicates which regions are permissible for magma loss (Fig. 2*b*, *c*). A relatively small region remains unforbidden. This is the region in which

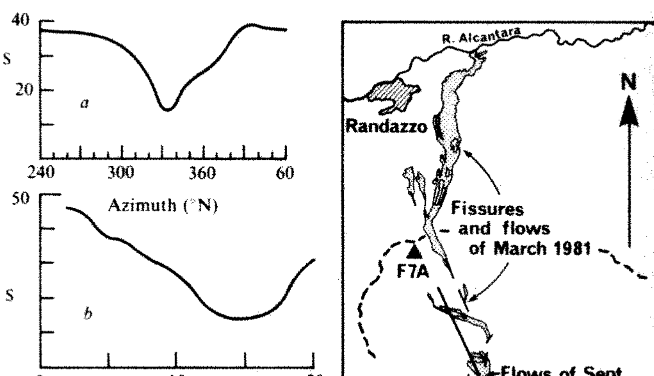


Fig. 3 Minimization of the statistic *S* against variation of azimuth (*a*), length (*b*) and depth (*c*) to upper surface of a dyke modelling the March 1981 eruption. The optimum dyke is shown in plan (*d*) and in section (*e*) relative to the eruptive fissures. (Map of fissures and flows from mapping by C. R. J. Kilburn (September 1980) and Guest *et al.* (March 1981).)

magma could have been stored before eruption, assuming that the observed changes were due to magma loss and that the loss was localized. The optimum point for magma loss can be found by minimizing residuals, and occurs at a point 2 km east and 4 km south of the summit. This model corresponds to the 60% probability level on a χ^2 test with 16 d.f., and therefore cannot be rejected. The fact that a point mass solution is possible, and that it lies so far south of the summit may be due partly to the strong southern bias of the stations available for survey.

Let us consider a more realistic model. We relax the point mass requirement and suppose instead a thin vertical dyke of variable size and position. Because no data are available for the northern summit region we assume the dyke to be symmetrical about the summit convergence of known historical fissure systems, which lies about 1.5 km east-southeast of the central crater (G. Kieffer, unpublished). The total mass of the dyke is set to 4.7×10^{10} kg and its vertical extent, which is not a sensitive parameter, is set to 1,000 m.

Modelling is performed using a computer program which models a dyke of any azimuth, length and depth as a lattice of point masses. For each model the sum of squared residuals, S, is calculated, and we search for an absolute minimum. A pronounced minimum exists, and its local behaviour is shown in Fig. 3*a–c*, where S is plotted against variation of each parameter. The optimum dyke is of azimuth $335°\pm5°$, length $14\,\text{km}\pm3$ and has its top surface at elevation $1{,}000\,\text{m}\pm500$. For this dyke, $S=14.1$, and it cannot be rejected at the 50% confidence level. Its position in plan and section is shown in Fig. 3*d*, *e*. It corresponds well with the surface activity of the March 1981 eruption.

A similar set of measurements was made immediately after the sub-terminal eruptions of September 1980. These occurred from the North-East Crater and were short-lived and highly explosive. The total mass of eruptive products is estimated at 2.8×10^{10} kg (C. R. J. Kilburn, personal communication).

No broad-scale gravity reduction was observed, and there were no reductions significant at the 95% level. Only one station, F9 to the north-east, showed a gravity reduction significant at the 68% level. Results are displayed, as before, in Table 1 and Fig. 1*c*. Stations on the south and south-east flanks showed a continuation of an upward trend which had been observed since 1979 and stations along the western flanks showed a pronounced increase. The increases were in the range 20–40 μGal. At the summit there was no significant change. In the absence of negative changes we cannot provide a simple model for the magma source; no model gives a fit which cannot be discounted at the 95% level. If the source is constrained to lie beneath the summit the statistic S improves as the source is lowered but never becomes acceptable. On this evidence the summit effusion of lava in September 1980 was not associated with a reservoir at detectable depth. It is proposed that the observed positive gravity changes were due to filling of radial dykes, associated with a sudden uprush of magma to the summit from a source well below sea level.

I conclude that these results support the view that Mount Etna underwent two distinct types of eruption during this period: a highly explosive summit eruption from a source well below sea level; and a massive flank eruption associated with draining from a radial dyke near sea level, which had become charged with magma during the preceding 6–12 months.

This work was funded by NERC. I thank D. R. Cowan, J. E. Guest, C. R. J. Kilburn, R. G. Mason and J. B. Murray for useful discussions and comments.

Received 28 July; accepted 10 December 1981.

1. Wadge, G. *Nature* **288**, 253–255 (1980).
2. Mogi, K. *Bull. Earthquake Res. Inst. Tokyo Univ.* **36**, 99–134 (1958).
3. Eaton, J. P. *Geophys. Monogr. 6, The Crust of the Pacific Basin* (eds MacDonald, G. A. & Kuno, H.) 13–29 (AGU Publ. No. 1035, 1962).
4. Fiske, R. S. & Kinoshita, W. T. *Science* **165**, 341–349 (1969).
5. Murray, J. B. & Guest, J. E. *U.K. Research on Mount Etna 1977–79*, 18–20 (Royal Society, London, 1980).
6. Wadge, G. *J. Volcan. geotherm. Res.* **1**, 237–264 (1976).
7. Washington, H. S., Aurousseau, M. & Keyes, M. G. *Am. J. Sci.* **12**, 379–408 (1926).
8. Tanguy, J.-C. *Bull. Soc. géol. Fr. Ser. 7*, **8**, 201–217 (1966).
9. Sharp, A. D. L., Davis, P. M. & Gray, F. *Nature* **287**, 587–591 (1981).
10. Rittmann, A. *Volcanoes and their Activity* (Interscience, London, 1962).
11. Rittmann, A. *Phil. Trans. R. Soc.* **A274**, 5–16 (1973).
12. Guest, J. E. & Duncan, A. M. *Nature* **290**, 584–586 (1981).
13. Cristofolini, R. *Phil. Trans. R. Soc.* **A274**, 17–35 (1973).
14. Romano, R. & Sturiale, C. *Phil. Trans. R. Soc.* **A274**, 37–43 (1973).
15. Sanderson, T. J. O. & Mason, R. G. *U.K. Research on Mount Etna, 1977–79*, 20–22 (Royal Society, London, 1980).
16. Cartwright, D. E. & Tayler, R. J. *Geophys. J. R. astr. Soc.* **23**, 45–74 (1971).
17. Cartwright, D. E. & Edden, A. C. *Geophys. J. R. astr. Soc.* **33**, 253–264 (1973).
18. Mohamed, I. thesis, Univ. London (1979).
19. Oliver, H. W., Robbins, S. L., Grannell, R. B., Alewine, R. W. & Biehler, S. *Bull. Calif. Div. Mines Geol.* **196**, 195–211 (1975).
20. Moore, J. G. *Nature* **282**, 250–253 (1979).
21. Blake, S. *Nature* **289**, 783–785 (1981).
22. Tanguy, J.-C. & Kieffer, G. *Bull. volcan.* **40-4**, 239–252 (1976/77).
23. Jachens, R. C. & Eaton, G. P. *J. Volcan. geotherm. Res.* **7**, 225–240 (1980).
24. Sanderson, T. J. O. & Mason, R. G. *Geophys. J. R. astr. Soc.* **65**, 283 (1981).
25. Romano, R. *EOS* **62**, 142–143, 477–478 (1981).

Active *in situ* disaggregation of oceanic crust and mantle on Gorringe Bank: analogy with ophiolitic massives

Yves LaGabrielle

GIS Océanologie et Géodynamique, U. B. O., avenue Le Gorgeu, 29200 Brest, France

Jean-Marie Auzende

Centre Océanologique de Bretagne, BP 337, 29273 Brest cedex, France

The *Cyagor II* cruise (May 1981) using submersible *Cyana SP 3000* on Gorringe Bank off south-west Portugal confirmed the scheme proposed at the end of *Cyagor I* cruise[1,2] that during the middle Cretaceous (110 Myr BP), Gorringe Bank, an old transform fault of Triassic–Liassic age, appears to have tilted to the north-east leading to the exposure of a section of oceanic mantle and crust. Due to its location at the boundary between European and African plates, Gorringe Bank suffered tectonic reactivation and uplifting during compressive and shearing-type events that occurred at the Cretaceous–Tertiary boundary and during the Tertiary (Eocene–Miocene). During the 12 dives of the *Cyagor II* cruise (Fig. 1), phenomena linked to the disintegration of serpentinites and gabbros in the deep-sea domain have been studied *in situ* with more detail than before[3]. Such processes have been suggested from studies on samples dredged or cored in oceans, or from observations in ophiolitic belts but direct observation had not enabled the magnitude of such phenomena to be measured. The observations on Gorringe Bank suggest that the processes described here are active in many oceanic sites. They help us to understand better the significance of many ophiolitic detritic beds more or less closely associated with pelagic sediments. We now compare the observations made at sea, especially on Gorringe Bank with similar observations in ophiolitic zones in mountain belts, which suggests that the ophiolite complexes have formed in a similar way to the breccias seen on the ocean floor.

Nine dives in areas where gabbros (Ormonde) and serpentinites (Gettysburg) outcrop revealed a very jagged deep-sea landscape made of small terraced cliffs. Each cliff, strongly fractured feeds a wide scree-covered slope. Fractures become more numerous at the top of the cliffs so that they are crowned by an unstable jumble of blocks. The size of the clasts released by the serpentine and gabbro disaggregation ranges from 1 mm to a few metres (Fig. 2).

Some large blocks of gabbro and serpentinite (between 1 and a few metres across) can be found a long way from any outcrop (100–1,000 m) where no other gabbroic or serpentinitic detritus occurs in the sediment. Under gabbro cliffs, two domains can be distinguished (Fig. 3): the first comprises 15°–40° slopes

which are scree-covered and composed of heaps without any apparent order—small blocks, cobbles, pebbles mixed with coarse sands are all present. The absence of pelagic sediments on this scree reveals the incessant activity to which they are subjected. The second domain comprising 5°–15° slopes is an area of transition between the scree-covered slopes and the shelves covered with pelagic sediments—sand-banks scarcely covered with small blocks lay there. Sometimes sands form rhomboidal ripples moving downwards. These sand-ripples were often observed lying on pelagic sediments. These sands were found to be composed of pyroxene, monomineralic fragments of plagioclase and, less frequently, of dolerite rock fragments.

The serpentinite cliff feeds extremely varied scree with both decimetric blocks and gravels. At the foot of the scree, pelagic mud is strewn with small blocks and unjointed broken stones.

The *in situ* observations made on Gorringe Bank include those of active alteration processes in the deep-sea environment, as well as of erosion and transport of ophiolitic materials. The processes that were seen can be linked to different causes. The first is the presence of rocky outcrops. The observations underline the multiplicity of escarpments, cliffs and fracture zones where gabbros and serpentinites are exposed. An intensive alteration is favoured by the internal fabric of the rocks themselves, particularly in the case of the gabbros. The alteration of granular type is linked with the crystallization of clays which are formed instead of plagioclases and then, release the grains of pyroxene. On the other hand, mechanical disaggregation takes place preferentially in the tectonized zone. In the case of Gorringe Massif, the known tectonic events[1,2] date from Triassic–Liassic when the transform motion deduced from plate geometry was recorded by the development of shear zones in the gabbros and serpentinites.

This shearing of the compressive tectonics which are still affecting the Massif, as shown by the level of natural seismicity in the area and by first-motions studies[16], has undoubtedly facilitated the creation of these highly tectonized zones. In the case of serpentinites, the tectonization plays a fundamental part in producing lenticular fracturing of the rock to give easily detached angular fragments of different sizes.

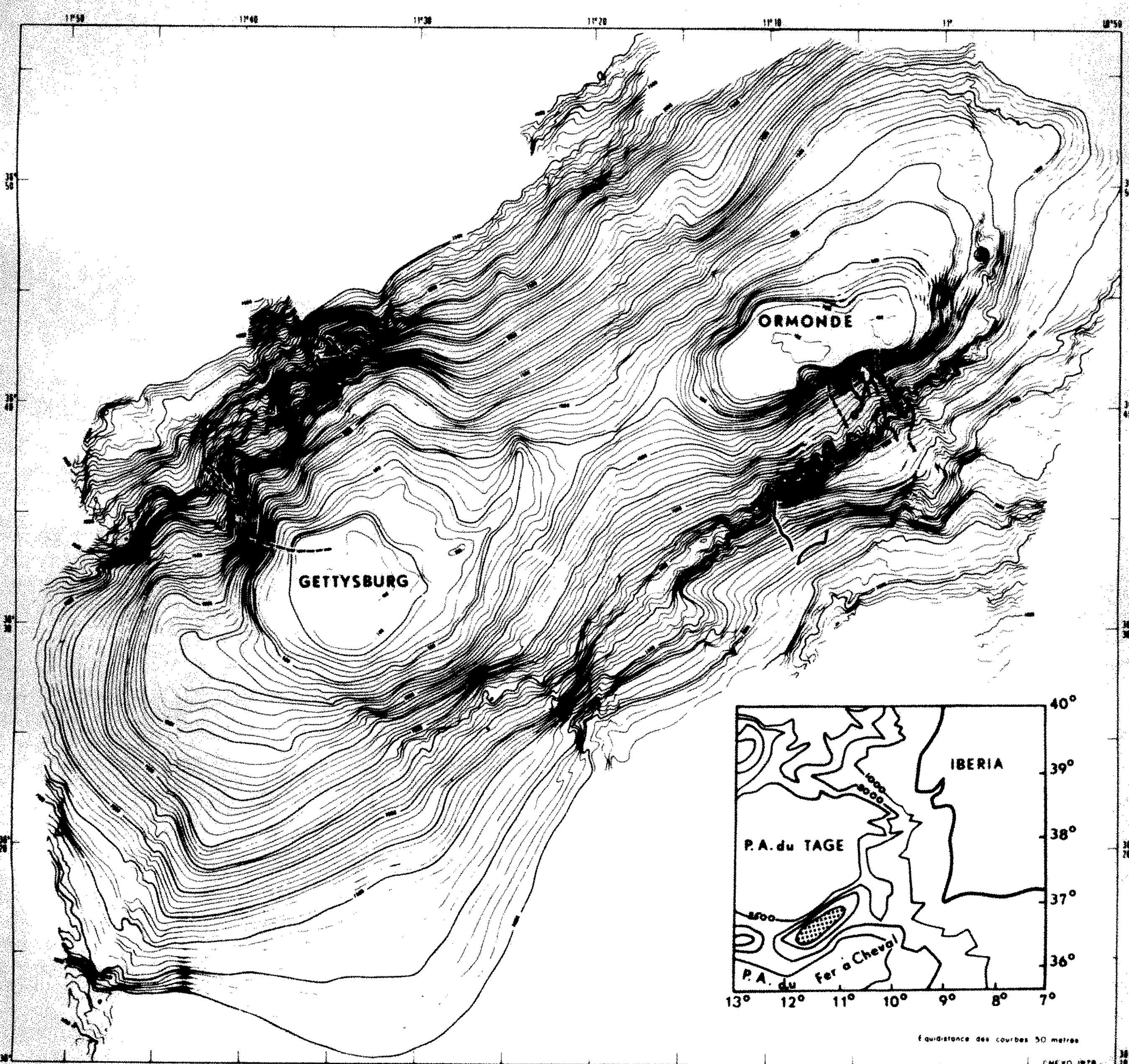


Fig. 1 Bathymetric map of Gorringe Bank with locations of *Cyana SP 3000* dives. Solid lines, *Cyagor II* cruise 1981; dashed line, *Cyagor I* cruise 1977.

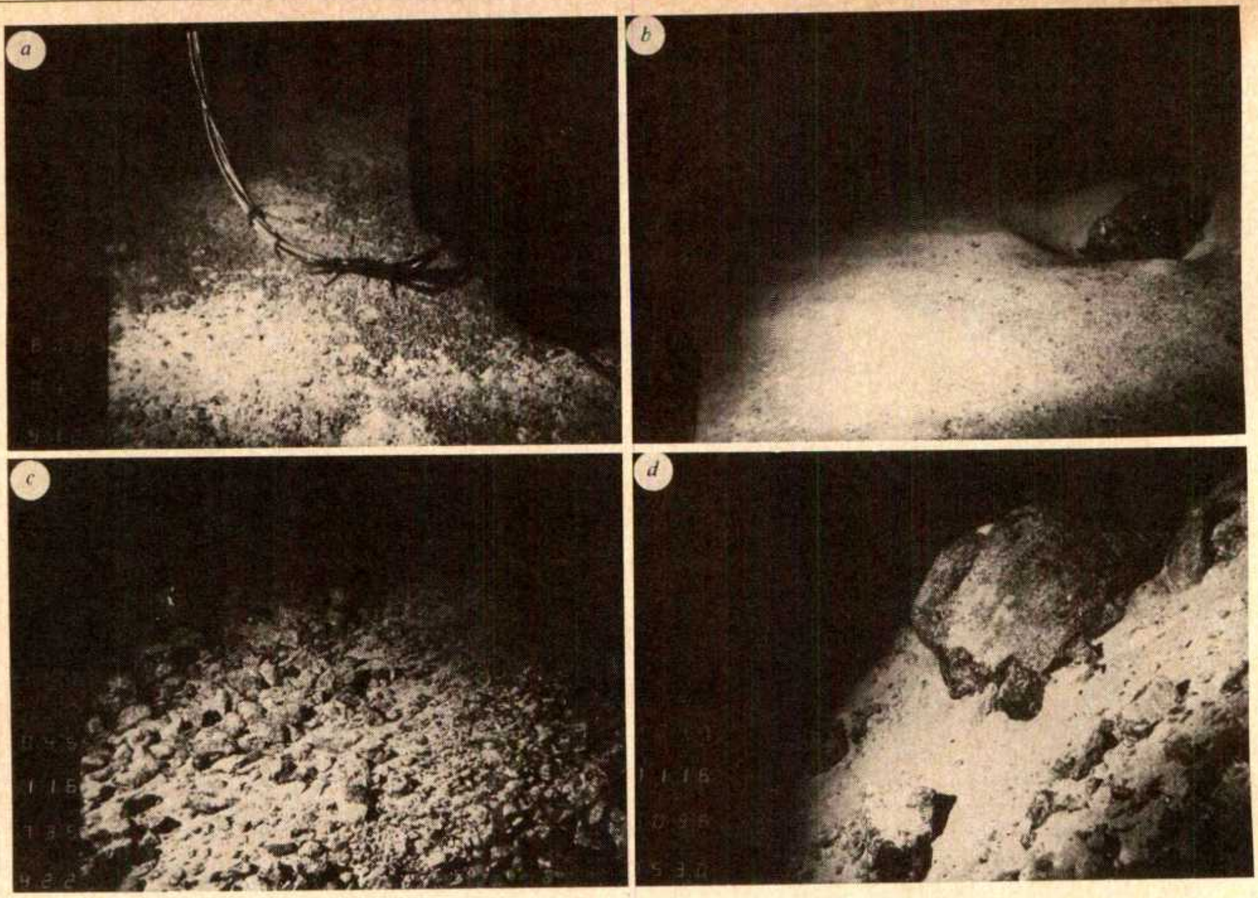


Fig. 2 Photographic plate: *a*, sands and gravels of gabbro; *b*, isolated block of gabbro; *c*, cobbles, pebbles and coarse sands of serpentinite; *d*, blocks of serpentinite in pelagic sediments.

The debris broken up by the double effect of granular alteration and tectonic disaggregation occupies a large area near the foot of the escarpment. Associated gravitional mass flows, distalites and coarse turbidites are bound to have participated in the filling of the adjacent abyssal plains.

Material, probably resulting from processes similar to those we have observed, has been located in many parts of the oceans, particularly in the proximity of complex structures involving a great variety of rocks of the mantle and of the oceanic crust such as mid-ocean ridges[4,11,12,17–20], fracture zones[7,8,21–33], and marginal basins[34–42]. These materials are likely to be more important than previously estimated in the sedimentary series of oceanic basins.

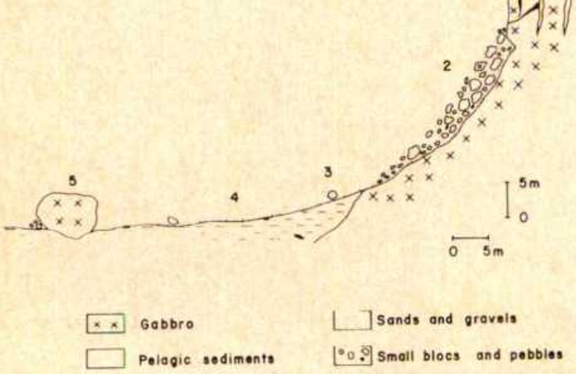


Fig. 3 *In situ* observations on Gorringe Bank (Mont Ormonde). 1, Highly fractured bedrock; 2, unbedded blocks, pebbles and sand at the foot of an escarpment; 3, small blocks isolated in sand and gravels; 4, sand and gravel ripples; 5, block of metric size.

Thus the genetic model obtained when studying oceanic data leads us to consider that any outcrop of the oceanic crust, whether tectonized or not, supplies detrital material. Most of this material remains *in situ* at the foot of the cliffs. The remainder is scattered in the surrounding autochthonous sedimentation mainly in two forms: metre-sized blocks which may slide down the slopes, and fine grained or coarse sand transported by turbidity currents. This latter fraction is consequently included in diversified and genetically different formations.

In mountain belts two types of deposits may be found for ophiolitic detrital products; *in situ*, that is to say closely associated with large igneous ophiolitic bodies, and scattered in the metasedimentary series[13–15,43–51]. These series may form independent tectonic units, so that the ophiolitic detrital material they contained will seem totally disconnected from its sources.

One can note a striking analogy between old and present day oceanic crust detritism found in the ocean and ophiolitic debris found in the mountain belts. However, we conclude that these types of detritic formations probably occupy a significant place in the total volume of ancient and present day oceanic sediments.

The results obtained in the ocean and particularly the *in situ* observations in the Gorringe Massif enable us to specify the morpho-structural framework of these present day and ancient deposits (cliffs, rugged escarpments, small basins). In all known cases in the oceans these structures belong or have belonged to areas topographically high in comparison with the oceanic floor (fracture zones, ridges, aseismic ridges). Thus, when these detritic levels are found in the chains, one must wonder if there is a relation between their genesis and the presence in the old ocean (Tethys Ocean, for example) of structures of abnormal topography. Those structures could result from an early intra

oceanic tectonic event (as on Gorringe, for example) and, despite variable causes, could be constant.

We thank J. F. Stephan, R. Blanchet and H. D. Needham for reviewing this article, the officers and crew of N. O. *Suroit*, the submersible *SP 3000 Cyana* team, and M. L. Augris for helping with the translation of the manuscript. Contribution no. 762 of the Centre Océanologique de Bretagne.

Received 19 November 1981; accepted 1 April 1982.

1. Auzende, J. M. *et al. Nature* **273,** 45–49 (1978).
2. Auzende, J. M. *et al. Bull Soc. géol. Fr.* **21,** 545–556 (1979).
3. LaGabrielle, Y. *et al. C.r. hebd. Séanc. Acad. Sci., Paris* **293,** série II, 827–832 (1981).
4. Fox, P. J. & Heezen, B. C. *Science* **149,** 1367–1370 (1965).
5. Lockwood, J. P. *Bull. geol. Soc. Am.* **82,** 919–936 (1971).
6. Lockwood, J. P. *Geol. Soc. Am. Mem.* **130,** 55–75 (1971).
7. Bonatti, E., Honnorez, J. & Gartner, S. *J. sedim. Petrol.* **43,** 728–735 (1973).
8. Bonatti, E., Emiliani, C., Ferra, G., Honnorez, J. & Rydell, H. *Mar. Geol.* **16,** 83–102 (1974).
9. Barett, T. J. & Spooner, E. T. C. *Earth planet. Sci. Lett.* **35,** 79–91 (1977).
10. Cortesogno, L., Galbiati, B., Principi, G. & Venturelli, G. *Ofioliti* **3,** 99–160. (1978).
11. Roberts, W. P. *Init. Rep. DSDP* Leg **49,** 421–426 (1978).
12. Varet, J. & Demange, J. *Init. Rep. DSDP* Leg **49,** 749–760. (1978).
13. Lemoine, M. & Tricart, P. *C.r. hebd Séanc. Acad. Sci. Paris.* **D288,** 1655–1659 (1979).
14. Lemoine, M. *Abstr. Symp. Genève* (1979).
15. Hebert, R. *Can. J. Earth Sci.* **18,** 619–623 (1981).
16. Udias, A., Lopez Arroyo, A. Mezcua, J. *Tectonophysics* **31,** 259–289 (1976).
17. Aumento, F. *et al. Init. Rep. DSDP* Leg **37** (1977).
18. Melson, W. G. *et al. Init. Rep. DSDP* Leg **45** (1978).
19. Dmitriev, L. *et al. Init. Rep. DSDP* Leg **46** (1978).
20. Luyendyk, B. P. *et al. Init. Rep. DSDP* Leg **49** (1978).
21. Olivet, J. L., Le Pichon, X., Monti, S. & Sichler, B. *J. geophys. Res.* **79,** 2059–2072 (1974).
22. Eittrem, S. & Ewing, J. *Geology* **3,** 555–558 (1975).
23. Fox, P. J., Schreiber, E., Rowlett, H. & McCamy, K. *J. geophys. Res.* **81,** 4117–4128 (1976).
24. Bonatti, E. *Earth planet. Sci. Lett.* **37,** 369–379 (1978).
25. Choukroune, P., Francheteau, J. & Le Pichon, X. *Bull. geol. Soc. Am.* **89,** 1013–1029 (1978).
26. Delong, S. E., Dewey, J. F. & Fox, P. J. *J. geogr. Soc. Lond.* **136,** 303–310 (1979).
27. Purdy, G. M., Rabinowitz, P. D. & Velterop, J. J. A. *Earth planet Sci. Lett.* **45,** 429–434 (1979).
28. Ludwig, W. J. & Rabinowitz, P. D. *Mar. Geol.* **35,** 99–110 (1980).
29. Hekinian, R. *et al. Mar. Geol.* (in the press).
30. Heezen, B. C., Bunce, E. T., Hersey, J. R. & Tharp, M. *Deep Sea Res.* **11,** 11–33 (1964).
31. Bonatti, E. *Nature* **219,** 363–364 (1968).
32. Bonatti, E., Honnorez, J. & Ferrara, G. *Phil. Trans. R. Soc.* **268,** 385–402 (1971).
33. Hekinian, R. & Aumento, F. *Mar. Geol.* **14,** 47–72 (1973).
34. Dickinson, W. R. *Soc. Econ. Paleont. Miner. Spec. Publ.* **19,** 230–239 (1974).
35. Karig, D E. & Moore, G. F. *Earth planet. Sci. Lett.* **26,** 233–238 (1975).
36. Blanchet, R. *Int. Symp. on the Structural History of the Mediterranean Basins,* 47–72 (Technip, Paris, 1977).
37. Sigurdsson, H., Sparks, R. S. J., Carey, S. N. & Haung, T. C. *J. Petrol.* **88,** 523–540 (1980).
38. Karig, D. E. *et al. Init. Rep. DSDP* Leg 31 (1975).
39. Blanchet, R. *et al. Bull. Soc. géol. Fr.* **21,** 529–532 (1979).
40. Gianelli, G. *Ofioliti* **2,** 115–135 (1977).
41. Simonian, K. O. & Gass, I. G. *Bull. geol. Soc. Am.* **89,** 1220–1230 (1978).
42. Ohnenstetter, M. *C.r. hebd. Séanc. Acad. Sci. Paris* **D289,** 1199–1202 (1979).
43. Stone, P. & Strachan, I. *Nature* **293,** 455–457 (1981).
44. Fudral, S., Rampnoux, J. P. & Robert, D. *C. somm. Soc. geol. Fr.* **6,** 330–332 (1977).
45. Lagabrielle, Y. *C.r. hebd. Séanc. Acad. Sci. Paris,* **292,** 1405–1408 (1981).
46. Elter, P. *Bull. Soc. géol. Fr.* **17,** 948–997 (1975).
47. Caby, R., Michard, A. & Tricart, P. *C.r. hebd. Séanc. Acad. Sci. Paris,* **D273,** 999–1002 (1971).
48. Folk, R. L. & MacBride, E. F. *J. sedim. Petrol.* **48,** 1069–1102 (1978).
49. Marini, M. & Terranova, R. *Ofioliti* **4,** 427–433 (1979).
50. De Wewer, P. & Caby, R. *C.r. hebd. Séanc. Acad. Sci. Paris* **292,** série II, 467–472 (1971).
51. LaGabrielle, Y., Nervo, R., Polino, R. & Dutto, F. *Abstr. Congr. Firenze* (1981).

Spherulitic crystallization in lamprophyric magmas and the origin of ocelli

Harald Carstens

Department of Geology, Aarhus University, DK-8000 Århus C, Denmark

Ocelli are rounded structures containing polycrystalline feldspar aggregates and minor amounts of mafic minerals; they occur in lamprophyre dykes all over the world. Ocelli are usually interpreted in terms of liquid immiscibility[1–3]. In alkaline lamprophyre dykes along the south coast of Norway[4], ocelli are closely associated with globular structures (varioles) which I have assumed[5] were formed as a result of a splitting of the magma into two fractions. The re-examination of these Norwegian dykes reported here, however, has shown that varioles, ocelli and leucocratic feldspar-rich segregations formed by spherulitic and closely related growth mechanisms from undercooled volatile-rich magmas without an intervening stage of liquid phase separation.

The morphology of the crystalline aggregates varies widely and is related to the distance from the dyke margins. The dykes nearly lie flat and never exceed 2 m in thickness. Depending on the mode of crystallization three different zones can be distinguished (Fig. 1). (1) A marginal zone, including a chill zone, characterized by globular growth units. Mutual impingement of the globules has here resulted in the formation of a polyhedral cell structure (Fig. 2). Dykes <30 cm thick are globular throughout. The thickness of the marginal zone in the larger dykes is 10–15 cm. The globules exhibit features characteristic of spherulitic growth. (2) An intermediate zone containing feldspatic segregations. This zone is only present in the upper part of dykes >50 cm thick. (3) A central homogeneous zone.

A very fine-grained chilled zone, 2–3 mm thick, is usually present in the marginal zone. The chilled zone contains many globules, the size of which is remarkably uniform and ~0.1 mm in diameter. In polarized light, each globule shows a black cross demonstrating that crystal fibres radiate in all directions from a central nucleation point. Microphenocrysts of kaersutite, titaniferous magnetite and augite are commonly enclosed in the globules. Kaersutite and augite are arranged with their long dimensions parallel to the dyke walls. The globules are usually arranged in long sub-parallel chains or layers parallel to the dyke margins presumably reflecting nucleation and growth along flow lines in the magma.

The size of the globules and the polyhedral cells, hereafter termed spherulites, increases away from the chilled zone and may become 10–12 mm in diameter 10–15 cm from the contact. The spherulites are essentially composed of alkali-feldspar, kaersutite, ±augite and Ti-magnetite. The feldspar forms elongate, curved and branching crystals radiating from a central point. However, with increasing grain size inwards from the dyke margins it often becomes impossible to recognize radiate crystal arrays because any section which does not intersect the centre of the spherulite appears to consist of a granular crystalline aggregate.

The polyhedral cells are separated by a layer of a greenish material, which sometimes contains pyrite. The thickness of this layer is <10 μm in the chill zone and increases to ~100 μm in the dyke marginal zone.

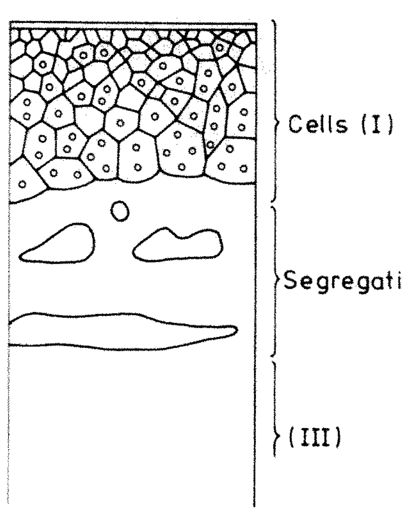


Fig. 1 Development of cells and leucocratic segregations in the marginal (I) and intermediate (II) zones respectively. Ocelli are indicated by circles. Not drawn to scale.

0028–0836/82/230493—02$01.00

Except for the finest compact spherulites of the chill zone, the spherulites always contain some rounded patches consisting almost exclusively of alkali feldspar. In each patch the crystals radiate from a common centre (Fig. 3*a*), and inwards in the dykes there is a change from a fibrous to a more lath-shaped morphology. These cone-shaped growth units (fan spherulites) apparently nucleated along the growth front of the larger spherulites in which they occur. It is frequently observed that they have grown from side-branches of radiating feldspar crystals (Fig. 3*b*). The fan spherulites actually represent the ocelli of the earlier description[4], which imposes constraints on the formation of the features. The observations preclude the possibility that the ocelli separated as liquid droplets.

With increasing distance from the dyke walls the polyhedral cell structure gradually disappears. This is due to a transition in morphology of the spherulites from compact to open, in the terminology of Keith and Padden[6]. When two open spherulites meet during growth the lamellae from both extend across the 'boundary' producing an interleaving structure.

The leucocratic segregations[4] in the intermediate zone are closely related to the ocelli of the marginal zone. The most abundant phase, alkali feldspar, shows a marked tendency to occur in thin bands or lenses roughly parallel to the dyke walls. Some bands consist of open, coarse spherulites, others consist of curved, branching feldspar crystals with preferred elongation inwards forming a comb-layer[7,8]. Kaersutite, which comes next in abundance, does not appear to grow directionally. The lamprophyre itself is very rich in kaersutite and Ti-magnetite along the upper and lower boundary of the segregations. The segregations contain scattered rounded ocellar structures, consisting of superimposed feldspar lamellae which are attached to one another like pages in an open book.

Chilled zones arise when the magma, coming in contact with the cold walls of the host rock, becomes greatly undercooled by the high rates of heat extraction. The higher the cooling rates, the lower the temperature at which nucleation occurs. Spherulitic growth is generally associated with rapid cooling rates[9], and chill zones would thus be an ideal environment for the growth of spherulites. The withdrawal of heat formed thermal gradients normal to the dyke walls, and crystallization proceeded from the margin inwards. The increasing size of the polyhedral cells away from the contact shows that the rate of nucleation decreased in the same direction. The growth of the spherulites approached a steady state and solidification terminated by the impingement of adjacent spherulites. The mafic sulphide-rich boundary layer which separates the cells has resulted simply from the impingement of enveloping boundary layers of rejected elements advancing ahead of the outer boundary of the spherulites. The thickness of the layer should be of the order of twice the characteristic distance of the boundary layer or $2\,D/G$ where D is the diffusion coefficient and G the growth rate[6]. Assuming, for example, a diffusion coefficient of $10^{-7}\ \mathrm{cm^2\ s^{-1}}$, the growth rate of the spherulites in the chill zone would be of the order of $0.1\ \mathrm{mm\ min^{-1}}$.

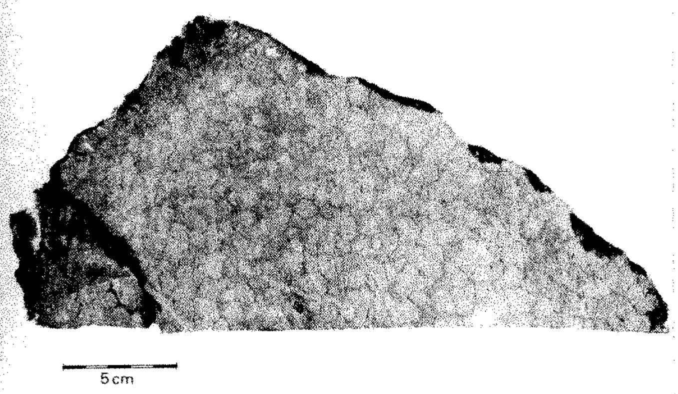


Fig. 2 Weathered specimen showing polyhedral cells. The structure is never observed on fresh surfaces, but may be developed by etching. The ocelli are 0.5–1 mm in diameter.

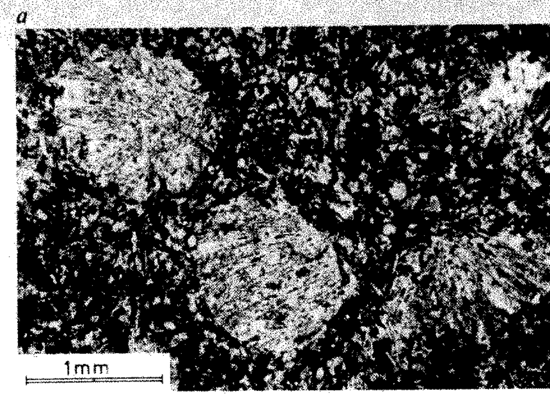


Fig. 3 *a*, Fan spherulites (ocelli), all belonging to the same cell. *b*, Branching feldspar terminating in fan spherulites (ocelli).

The textural developments seem to be related to feldspar nucleation difficulties in a lamprophyric magma, and the crystallization sequence is considered to have been determined by the relative ease of nucleation of the minerals involved rather than by equilibrium phase relationships. Delayed nucleation of feldspar is common in systems varying in composition from granitic to basaltic[9] presumably because of a slow rate of formation of embryonic clusters of equilibrium size. Thus, feldspar phenocrysts are always absent in lamprophyres and the fundamental shape of the feldspar is frequently spherulitic/dendritic indicating crystallization from an undercooled liquid. In the marginal zone, the high rate of cooling produced undercoolings of sufficient magnitude to permit the nucleation of feldspar simultaneously with the mafic minerals with the result that spherulites of melt composition grew and formed a cell structure. In the intermediate zone, slower cooling rates induced early nucleation of the mafic minerals relative to feldspar. The growth of kaersutite and magnetite established composition gradients in the melt perpendicular to the dyke walls. The melt was exposed to an increasing degree of undercooling by loss of water and iron to crystallizing kaersutite and ore minerals, and when feldspar finally did nucleate, conditions were favourable for the growth of feldspar-rich spherulites or alternatively inward growth of fan-shaped bundles of feldspar. The leucocratic segregations thus formed by periodic build-up of appropriate conditions of undercooling.

Regardless of the detailed mechanisms involved, the present study has shown that the globules (varioles), the polyhedral cells, the leucocratic segregations, and the ocelli which occur in them, are polycrystalline aggregates formed by crystal growth, the structural variations being a kinetic effect related to cooling rates. The observations presented here have produced no evidence in favour of liquid–liquid phase separation, although this still remains a possible explanation for the formation of other ocellar structures.

I thank Gary C. Lofgren for valuable discussions and criticism and C. H. Donaldson for helpful comments.

Received 15 February; accepted 19 April 1982.

1. Philpotts, A. R. *Am. J. Sci.* **276**, 1147–1177 (1976).
2. Eby, G. N. *Contr. Miner. Petrol.* **5**, 269–278 (1980).
3. Freestone, I. C. *Chem. Geol.* **23**, 115–123 (1978).
4. Carstens, H. *Beitr. Miner. Petrol.* **6**, 299–319 (1959).
5. Carstens, H. *Chem. Geol.* **27**, 297–307 (1979).
6. Keith, H. D. & Padden, F. J. *J. appl. Phys.* **34**, 2409–2421 (1963).
7. Lofgren, G. E. & Donaldson, C. H. *Contr. Miner. Petrol.* **49**, 309–319 (1975).
8. Donaldson, C. H. *Miner. Mag.* **41**, 323–336 (1977).
9. Lofgren, G. in *Physics of Magmatic Processes* (ed. Hargraves, R. B.) 487–551 (Princeton University Press, 1980).

Isolation of a natural hatching stimulus, glycinoeclepin A, for the soybean cyst nematode

Tadashi Masamune, Masaki Anetai, Mitsuo Takasugi & Nobukatsu Katsui

Department of Chemistry, Faculty of Science, Hokkaido University, Sapporo 060, Japan

The cyst nematodes (genera *Heterodera*, *Globodera* and others)[1] generally have a limited host range. This specificity arises from the fact that the hatching of larvae from cysts occurs in response to specific stimuli secreted by the host plants[2–4]. Because of the scientific interest and importance in agriculture of these stimuli, since their first recognition[5] in 1922 much time and effort[2–4] has been expended in attempts to purify and characterize them. However, success has been limited by the extremely small amounts of active substance available. We report here the isolation of a natural hatching stimulus for the soybean cyst nematode from the roots of kidney bean. The substance, designated as glycinoeclepin A, stimulates the hatching of larvae from the eggs *in vitro* at a concentration of 10^{-11} to $\sim 10^{-12}$ g ml^{-1} in water at 25 °C.

The soybean cyst nematode (*Heterodera glycines* Ichinohe) parasitizes a small group of host plants, including soybean (*Glycine max*), kidney bean (*Phaseolus vulgaris*) and adzuki bean (*Phaseolus angularis*), and causes 'soybean sickness'[6]. Tsutsumi and Sakurai[7] demonstrated that the hatching of larvae from cysts of the nematode was stimulated by root diffusates from the host plants. Since then various factors affecting hatching have been examined in detail[8–11]. The present investigation is based on these findings.

The bioassay for hatching was performed with free eggs. The cysts were isolated from infected soil collected in October at Memuro, Hokkaido, and stored in a refrigerator at 5 °C. Before the test, the cysts were soaked in water at 25 °C for 10 days, then dissected in water and sieved through a 325-mesh sieve. The dissected cyst wall fragments remained on the sieve. The eggs and larvae were transferred to a beaker of water so that 1 ml of the aqueous suspension contained 150–200 eggs. Then 1 ml of the suspension was transferred to a Syracuse watch glass, to which 8 ml of distilled water and 1 ml of the test solution were added. The whole suspension was kept at 25 °C and the number of the larvae hatched was counted after 10 days. The per cent hatching rate was expressed as $P = P_t - P_w$, where P_t and P_w denote the per cent hatching observed in the test solution and in distilled water, respectively. The test solution was said to be active when the hatching rate, P, exceeded 50%. Hatching rate is an adequate measure of hatching and emergence[7].

Hatch-stimulating concentrates were prepared from aqueous extracts of the roots of two subspecies ('Hon-kintoki' and 'Beni-kintoki') of kidney bean rather than from their root diffusates. Dried and powdered roots (113 kg) of Hon-kintoki, collected in late July at Memuro (1 hectare), were extracted with water at a temperature of <10 °C, and the aqueous suspension, after concentration at <30 °C, was acidified to *p*H 2–3 and extracted again, with chloroform. The chloroform solution gave 'chloroform extracts A' (102 g), which were fractionated according to the procedure summarized in Table 1, guided by the bioassay. Repeated chromatography yielded a highly active fraction, fraction L (~50 μg). Dried roots of Beni-kintoki were also treated and fractionated in almost the same manner to give the corresponding fraction with the same activity. These fractions were homogeneous and identical by analytical HPLC over various columns (μ-Bondapak NH_2, MicroPak NH_2-10, μ-Bondapak C_{18} and μ-Bondapak/Phenyl) and by the respective field desorption (FD) mass spectra (Fig. 1). The identification of glycinoeclepin A as the active substance was also checked by the division of fraction L into five fractions by chromatography on two separate columns (μ-Bondapak NH_2 and μ-Bondapak C_{18}) followed by the bioassay of each fraction after hydrolysis.

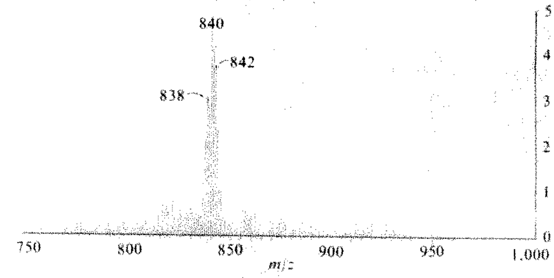


Fig. 1 FD mass spectrum of *p*-BPE of glycinoeclepin A.

The FD mass spectrum (Fig. 1) established that the bis(*p*-bromophenacyl) ester of glycinoeclepin A (GEA *p*-BPE) has a molecular weight of 838 and hence that glycinoeclepin A is a dibasic acid with the molecular weight of 446. The high-resolution FD mass spectrum ($m/z = 838.1334$) indicated that GEA *p*-BPE is best represented by $C_{41}H_{44}O_9Br_2$ (calculated $m/z = 838.1352$) and hence the acid by $C_{25}H_{34}O_7$. This was supported by the ^{1}H-NMR spectrum (Fig. 2) of GEA *p*-BPE, which revealed about 34 hydrogen and 7 oxygen atoms (two —COOH, one —OH, one —O— and one $-\overset{|}{C}{=}O$) in the acid.

Table 1 Isolation of glycinoeclepin A as its bis(*p*-bromophenacyl) ester

Dried and powdered roots of kidney bean 'Hon-kintoki' (113 kg)
↓ in the text
Chloroform extracts A (102 g, active at 10^{-5} g ml^{-1})
↓ EtOAc–aq.$NaHCO_3$
Acidic extracts B (57 g, active at 10^{-5} to $\sim 10^{-6}$ g ml^{-1})
↓ Charcoal–Celite column (aq. acetone)
Fraction C (21.5 g, active at 10^{-6} g ml^{-1})
↓ Sephadex LH-20 column (MeOH)
Fraction D (10.1 g, active at 10^{-6} to $\sim 10^{-7}$ g ml^{-1})
↓ DEAP–Sephadex LH-20 column (72% EtOH–AcOH–NH_4OH)
Fraction E (2.16 g, active at 10^{-7} g ml^{-1})
↓ Sephadex LH-20 column (CH_2Cl_2–MeOH, 95:5)
Fraction F (135 mg, active at 10^{-8} g ml^{-1})
↓ *p*-Bromophenacylation (*p*-$BrC_6H_4COCH_2Br$ and (i-Pr)$_2$NEt in MeCN, room temperature, overnight) and silica gel column
Fraction G as *p*-BPE (316 mg, active at 10^{-8} g ml^{-1} after hydrolysis)
↓ Preparative HPLC, Hitachi Gel 3019 column (MeOH–CH_2Cl_2, 7:3)
Fraction H as *p*-BPE (112 mg, active at 10^{-8} to 10^{-9} g ml^{-1} after hydrolysis)
↓ Preparative HPLC, Bondapak C_{18}/Porasil B (MeOH–H_2O, 8:2)
Fraction I as *p*-BPE (29 mg, active at 10^{-9} g ml^{-1} after hydrolysis)
↓ Preparative HPLC, μ-Bondapak NH_2 column (C_6H_{14}–CH_2Cl_2–MeCN, 70:18:12)
Fraction J as *p*-BPE (2.2 mg, active at 10^{-10} g ml^{-1} after hydrolysis)
↓ Preparative HPLC, μ-Bondapak NH_2 column (C_6H_{14}–CH_2Cl_2–MeCN, 80:12:8)
Fraction K as *p*-BPE (1.2 mg, active at 10^{-10} g ml^{-1} after hydrolysis)
↓ Preparative HPLC, μ-Bondapak NH_2 column (C_6H_{14}–CH_2Cl_2–MeCN, 60:36:4)
Fraction L (glycinoeclepin A) as *p*-BPE (~50 μg, active at 10^{-11} to $\sim 10^{-12}$ g ml^{-1} after hydrolysis)

0028-0836/82/230495—04$01.00

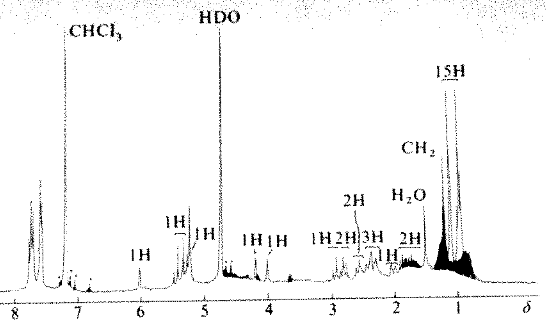


Fig. 2 ^{1}H-NMR spectrum of *p*-BPE of glycinoeclepin A (270 MHz, 50 μg in $CDCl_3 + D_2O$). 12,260 scans.

We thank Drs M. Tsutsumi, H. Inagaki and T. Okada for helpful discussions of the bioassay, Mr S. Ueno for helping to prepare extracts A, and Drs T. Higashijima and T. Higuchi for measuring NMR and mass spectra, respectively. The research was supported by a grant-in-aid for Special Project Research no. 56104009 provided by the Ministry of Education, Science and Culture of Japan.

Received 1 February; accepted 29 March 1982.

1. Stone, A. R. *Nematologica* **23**, 273–288 (1977).
2. Clarke, A. J. & Perry, R. N. *Nematologica* **23**, 350–368 (1977).
3. Shepherd, A. M. & Clarke, A. J. in *Plant Parasitic Nematodes* Vol. 2 (eds Zuckerman, B. M., Mai, W. F. & Rohde, R. A.) 267–287 (Academic, New York, 1971).
4. Okada, T. *Jap. J. Nematol.* **5**, 1–9 (1975).
5. Baunacke, W. E. *Arb. biol. Bund Anst. Land-u. Forstw.* **11**, 185–288 (1922).
6. Skotland, C. B. *Phytopathology* **47**, 623–625 (1957).
7. Tsutsumi, M. & Sakurai, K. *Jap. J. appl. Ent. Zool.* **10**, 129–137 (1966).
8. Okada, T. *Bull. agric. Chemicals Inspection Stn* **10**, 73–75 (1970).
9. Okada, T. *Appl. Ent. Zool.* **7**, 99–102, 234–237 (1972).
10. Okada, T. *Jap. J. appl. Ent. Zool.* **15**, 215–221 (1971).
11. Lehman, P. S., Barker, K. R. & Huisingh, D. *Nematologica* **17**, 467–473 (1971).

Long-term potentiation of the perforant path *in vivo* is associated with increased glutamate release

A. C. Dolphin, M. L. Errington & T. V. P. Bliss

Division of Neurophysiology & Neuropharmacology, National Institute for Medical Research, Mill Hill, London NW7 1AA, UK

Long-term potentiation (LTP) of synaptic transmission following brief trains of high-frequency stimulation in hippocampal pathways has attracted attention as a possible physiological correlate of memory. The mechanism of LTP is little understood, and what evidence there is suggests that both pre- and postsynaptic mechanisms are involved[1–7]. Previous findings that Ca^{2+} is required for LTP[8,9], and that LTP is accompanied by a prolonged increased in Ca^{2+} uptake[10], suggest that an increase in impulse-dependent release of transmitter, a process critically dependent on presynaptic Ca^{2+} levels[11,12], may be an important component of this form of synaptic plasticity. A recent report[13] that release of previously accumulated ^{3}H-D-aspartate is increased following tetanic stimulation of afferent fibres to CA1 pyramidal cells in hippocampal slices is also consistent with a presynaptic mechanism. However, there has been no direct evidence relating LTP to an increased stimulus-dependent release of transmitter. We have now investigated whether LTP in the perforant path (PP) of the rat is associated with a change in the release of neurotransmitter. A modification of the push–pull cannula technique has enabled us to infuse ^{3}H-glutamine and measure the subsequent release of newly synthesized ^{3}H-glutamate, and at the same time to monitor the field potentials evoked in the dentate gyrus by stimulation of the PP. We have found a prolonged increase in the release of glutamate, the probable transmitter of the PP[14,15], following the induction of LTP in this pathway.

In anaesthetized rats, a push–pull cannula was placed in the molecular layer of the dentate gyrus, where PP fibres terminate on the apical dendrites of granule cells. Two insulated stainless-steel wire recording electrodes were attached to the outer cannula so that the exposed tip of the upper electrode lay roughly midway between the ends of the inner and outer cannulae, and the lower electrode extended 0.2–0.4 mm beyond the inner cannula (Fig. 1*a*). Correct placement of the electrode/cannula assembly was achieved by lowering it to a depth at which the negative-going potential evoked by stimulation of the PP was maximal, as recorded through the upper (dendritic) electrode. The potential recorded through the lower (somatic) electrode, positioned in the granule cell body layer, was used to monitor synaptic transmission during the experiment; the amplitude of the initial positive-going component of this response, which we refer to as the synaptic wave (Fig. 1*a*), was used as a measure of the strength of synaptic activation[16]. Following the induction of LTP by a high-frequency stimulus train (see Fig. 1 legend) an increase in the amplitude of the synaptic wave normally occurs (Fig. 1*a*), persisting with little decrement after the first 15 min (Fig. 1*b*, high-frequency train delivered at the time indicated by the arrow, mean of 11 animals showing LTP).

To investigate the release of glutamate in the dentate gyrus, the area to be perfused was preloaded with the radiolabelled precursor, ^{3}H-glutamine. Because the enzyme glutaminase, which converts glutamine to glutamate, is concentrated in presynaptic terminals[17], most of the infused ^{3}H-glutamine is probably taken up by the glutamate-utilizing terminals of the PP, rather than in postsynaptic or non-neuronal elements[18–20]. ^{3}H-glutamine (20 μCi) was infused through the inner cannula over 30 min. One hour later, perfusion with oxygenated medium was begun. The ^{3}H-glutamate present in sequential fractions of the perfusate was separated and measured as described in Fig. 2 legend.

In control experiments, the release of ^{3}H-glutamate into the perfusate while the PP was stimulated continuously at 0.1 Hz showed a steady decline during the perfusion (Fig. 2*a*). The release of ^{3}H-glutamate was at least partially stimulus dependent, since in seven animals, when the rate of stimulation was increased 30-fold, from 0.1 to 3 Hz, and the stimulus voltage was doubled, 120 min after the start of perfusion, the release of ^{3}H-glutamate was markedly increased (Fig. 2*b*). Note, however, that although larger potentials were recorded with no decrement throughout this 75-min period of increased stimulation, the mean outflow of ^{3}H-glutamate returned to a level not significantly greater than control values within 45 min, possibly due to depletion of the pool of ^{3}H-glutamate.

In another set of animals, the pattern of ^{3}H-glutamate release following the induction of LTP was investigated. The PP was stimulated at 0.1 Hz throughout the experiment, and superimposed on this, a high-frequency train (250 Hz for 0.5 s, stimulus voltage doubled) was delivered to the PP 120 min after the start of perfusion (arrows in Figs 1*b*, 2*c* and *d*). LTP of the PP–granule cell synapse (as judged by an increase in the amplitude of the synaptic wave lasting for 30 min or more) was produced in 11 out of 17 animals. The mean time course of potentiation is shown in Fig. 1*b* for all the animals in which LTP occurred. All these animals showed a prolonged increase in the release of ^{3}H-glutamate after the induction of LTP; indeed, the average level of released ^{3}H-glutamate in the potentiated animals remained above the corresponding average control value until the end of the experiment, 75 min later (Fig. 2*d*). In contrast, the six animals which received a high-frequency train, but in which the criteria for LTP were not met, showed only a slight elevation of ^{3}H-glutamate release following the train compared with control animals (Fig. 2*c*). As a further control that the prolonged increase in ^{3}H-glutamate release shown in Fig. 2*d* was associated with the events following the high-frequency conditioning train rather than the train itself, the same number and strength of stimuli used to produce LTP were given at a lower frequency (2 Hz) to five animals, 120 min

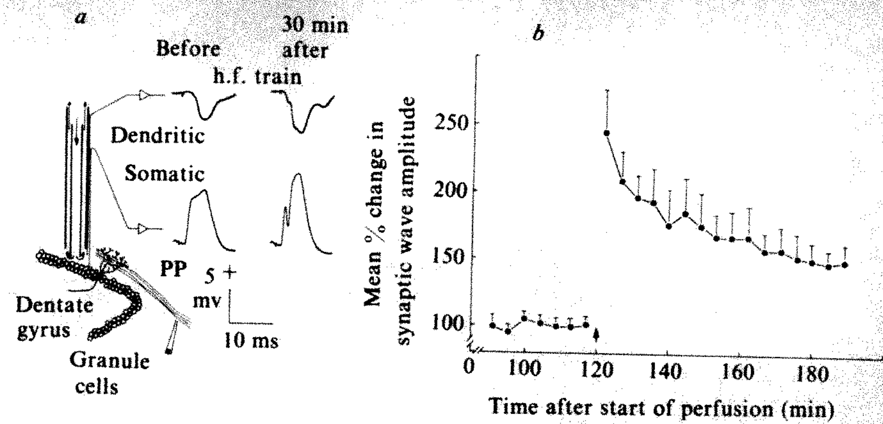


Fig. 1 Long-term potentiation in the perforant path during perfusion of the dentate gyrus through a push-pull cannula. Experiments were performed on male Sprague-Dawley rats (300–330 g), anaesthetized with urethane (1.5 g kg^{-1}, intraperitoneally). *a*, Two recording electrodes (insulated stainless-steel wire, 0.1 mm o.d.) were glued to the outer ('pull') tube of the concentric push-pull cannula (0.83 mm o.d., Plastic Products, Clark Electromedical). The upper electrode protruded 0.2 mm, and the lower electrode 0.5–0.7 mm below the outer cannula. The electrode/cannula assembly was lowered stereotaxically into the dentate gyrus of the hippocampus. The outer cannula was fixed to the skull with dental cement, and the inner ('push') cannula was then lowered so that its tip protruded 0.3 mm below the outer cannula. The PP was stimulated through a concentric bipolar stimulating electrode (Rhodes, Clark Electrochemical) with constant voltage pulses (0.05 ms, range 10–50 V) at a strength near threshold for a population spike (the negative-going spike superimposed on the synaptic wave and reflecting the discharge of granule cells). Throughout the experiment, evoked potentials were recorded every 90 s on a chart recorder, and on disc for subsequent computer analysis. Examples are shown of potentials recorded from the upper (dendritic) and lower (somatic) electrodes during perfusion in a single animal, immediately before and 30 min after a high-frequency (h.f.) stimulus train (250 Hz, 0.5 s, stimulus voltage doubled). *b*, Graph showing the mean percentage increase (±1 s.e.m.) in the amplitude of the synaptic wave for all animals ($n = 11$) which displayed LTP after receiving a high-frequency conditioning train 120 min after the onset of perfusion. In each animal the amplitude of the synaptic wave was measured at a fixed latency and expressed as a percentage of its mean amplitude during the 30 min preceding the conditioning train.

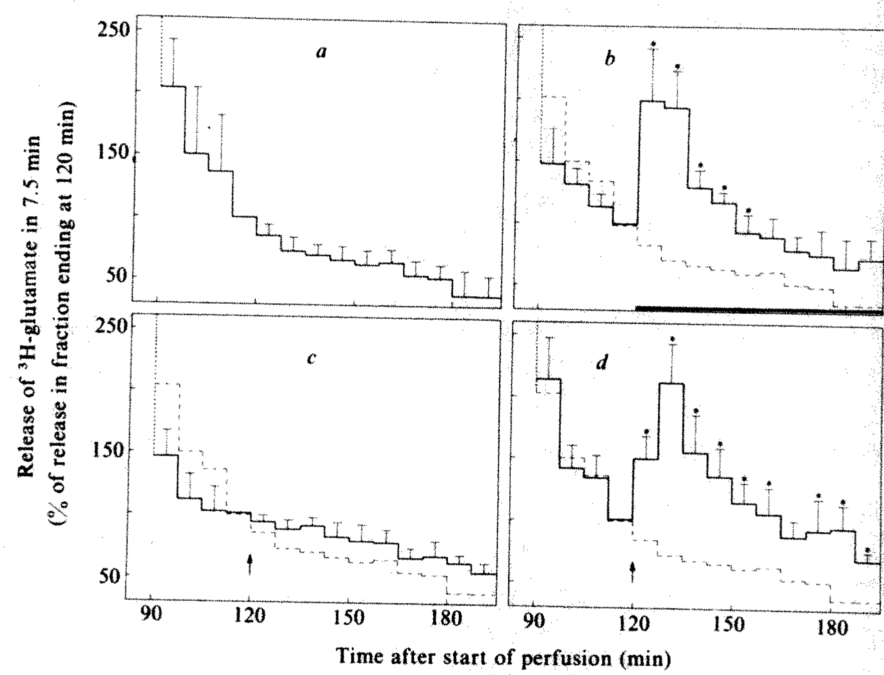


Fig. 2 *In vivo* release of ^{3}H-glutamate from the dentate gyrus perfused through a push–pull cannula. In all experiments ^{3}H-glutamine (20 µCi, 0.75 nmol) was infused in 3 µl oxygenated (95% O_2/5% CO_2) perfusion medium (composition in mM: Na^+, 152; K^+, 3.7; Mg^{2+}, 1.2; Ca^{2+}, 1.8; Cl^-, 140; $H_2PO_4^-$, 1.2; SO_4^{2-}, 1.2; HCO_3^-, 16; glucose, 10) containing a trace of phenol red, at a rate of 0.1 µl min^{-1}. Perfusion with continuously oxygenated medium at 10 µl min^{-1} was begun 1 h later, using a dual-channel peristaltic pump. Sequential 7.5-min fractions of perfusate were collected on dry ice. The phenol red served to mark the entry of the perfusate into the fraction collector, and thus the lag time of the apparatus (8 min). ^{3}H-glutamate was separated from ^{3}H-γ-aminobutyric acid and from their precursor ^{3}H-glutamine as previously described[21,22] using anion-exchange columns. The ^{3}H-glutamate released in a single experiment was expressed for each fraction as a percentage of the release in the fraction ending at 120 min. Values for each group of animals are given as mean ± 1 s.e.m. for all fractions between 90 and 195 min following the start of perfusion. *a*, Control perfusion, PP stimulated continuously at 0.1 Hz ($n = 7$). Values for control ^{3}H-glutamate release are repeated in *b*, *c* and *d* (dashed lines), where * indicates significant increase in ^{3}H-glutamate release compared with control ($P < 0.05$, Student's *t*-test). *b*, PP stimulated at 0.1 Hz until 120 min after start of perfusion when the rate was increased to 3 Hz and the stimulus strength was doubled for the remainder of the experiment as indicated by the solid bar ($n = 7$). *c*, *d*, PP stimulated continuously at 0.1 Hz, and at 120 min after the start of perfusion (arrow) a conditioning train was given (see Fig. 1 legend). *c*, ^{3}H-glutamate release in animals not showing LTP ($n = 6$). *d*, ^{3}H-glutamate release in animals showing LTP ($n = 11$).

after the start of perfusion. No potentiation occurred, and there was no difference in ^{3}H-glutamate release relative to controls, apart from a small increase in the fraction corresponding to the low-frequency train (results not shown).

Much evidence points to glutamate as having a transmitter role in the PP[14,15]. Our finding of a stimulus-dependent release of glutamate in this pathway (Fig. 2*b*) provides additional evidence for such a role. The main finding here, that LTP is associated with a prolonged increase in ^{3}H-glutamate release thus strongly suggests that a sustained increase in stimulus-dependent release of this transmitter is at least one of the mechanisms underlying LTP. This does not rule out the possibility that other changes involving, for instance, increased dendritic uptake of Ca^{2+} (ref. 10), altered dendritic morphology[4], or other mechanisms leading to an enhanced postsynaptic sensitivity[5] also contribute to LTP. The relative roles of pre- and postsynaptic events in the long-term enhancement of synaptic efficacy remain to be determined.

Received 19 February; accepted 21 March 1982.

1. Bliss, T. V. P. & Lømo, T. *J. Physiol., Lond.* **232**, 331–356 (1973).
2. Andersen, P., Sundberg, S. H., Sveen, O., Swann, J. W. & Wigström, H. *J. Physiol., Lond.* **302**, 463–482 (1980).
3. Dunwiddie, T. & Lynch, G. *J. Physiol., Lond.* **276**, 353–367 (1978).
4. van Harreveld, A. & Fifková, E. *Expl Neurol.* **49**, 736–749 (1975).
5. Baudry, M., Bundman, M. C., Smith, E. K. & Lynch, G. S. *Science* **212**, 937–938 (1981).
6. Browning, M., Bennett, W. F., Kelly, P. & Lynch, G. *Brain Res.* **218**, 255–266 (1981).
7. McNaughton, B. L., Douglas, R. M. & Goddard, G. V. *Brain Res.* **157**, 277–293 (1978).
8. Dunwiddie, T., Madison, D. & Lynch, G. *Brain Res.* **150**, 413–417.
9. Wigström, H., Swann, J. W. & Andersen, P. *Acta physiol. scand.* **105**, 126–128 (1979).
10. Baimbridge, K. G. & Miller, J. J. *Brain Res.* **221**, 299–305 (1981).
11. Dodge, F. A. & Rahamimoff, R. *J. Physiol., Lond.* **193**, 419–432 (1967).
12. Katz, B. & Miledi, R. *J. Physiol., Lond.* **216**, 503–512 (1971).
13. Skrede, K. K. & Malthe-Sørenssen, D. *Brain Res.* **208**, 436–441 (1981).
14. Storm-Mathisen, J. *Adv. Biochem. Psychopharmac.* **27**, 43–53 (1981).
15. Nadler, J. V. & Smith, E. M. *Neurosci. Lett.* **25**, 275–280 (1981).

16. Lømo, T. *Expl Brain Res.* **12,** 18–45 (1971).
17. Bradford, H. & Ward, H. K. *Brain Res.* **110,** 115–125 (1976).
18. Bradford, H., Ward, H. K. & Thomas, A. J. *J. Neurochem.* **30,** 1453–1459 (1978).
19. Hamberger, A., Chiang, G. H., Sandoval, M. E. & Cotman, C. W. *Brain Res.* **168,** 513–530 (1979).
20. Reubi, J.-C., Van der Berg, C. & Cuénod, M. *Neurosci. Lett.* **10,** 171–174 (1978).
21. Dolphin, A. C. *Brain Res.* (in the press).
22. Reubi, J.-C. *Neuroscience* **5,** 2145–2150 (1980).

Reversal of current through calcium channels in dialysed single heart cells

K. S. Lee & R. W. Tsien

Department of Physiology, Yale University, School of Medicine, New Haven, Connecticut 06510, USA

Calcium channels contribute to important cellular functions such as impulse conduction, rhythmic activity, muscle contraction, and secretion[1] and are a major target for modulation by neurohormones and drugs[2,3]. Despite their importance, Ca channels are much less well understood than Na and K channels because of several problems reviewed by Hagiwara and Byerly[1]. First, few preparations allow good control of electrical potential and ionic composition on both sides of the membrane, yet display strong Ca currents. Second, current carried by Ca channels is often obscured by ion movements through other membrane pathways. Third, outward current flow through Ca channels, carried by Ca^{2+} or any other ion, has been difficult to demonstrate. This is unfortunate because measurements of the potential at which current reverses (E_{rev}) have been crucial in understanding ion permeation and selectivity in other ionic channels (see ref. 4). In molluscan neurones, attempts at recording outward current through the Ca channel have not succeeded, largely because of overlap by nonspecific outward current[5–7]. In multicellular cardiac muscle preparations, strong depolarizations produce a decaying outward current that Reuter and Scholz attributed to K efflux through the Ca channel[8]. Their interpretation is controversial, however, since the evidence leaves open other explanations for the outward current—for example, Na efflux via Na–Ca exchange[9], or K efflux through K-selective channels. To overcome these problems, we studied Ca channels in single isolated heart muscle cells[10–12] using a suction pipette method[13]. We were able to record robust Ca currents with minimal interference from other time-dependent currents while controlling potential and ion composition on both sides of the membrane. Here we present experimental evidence for a genuine reversal of ionic current through Ca channels due to outward movement of K^+ ions, in support of the hypothesis of Reuter and Scholz.

Single isolated cells were obtained from the ventricles of adult guinea pig hearts by enzymatic dissociation using collagenase and hyaluronidase. Relaxed cells with cylindrical geometry were chosen for study; in a separate series of microelectrode recordings, such cells displayed resting potentials near −80 mV, vigorous contractions, and action potentials with full plateaus. A suction pipette (tip diameter 5–10 μm) provided access to the inside of the cell for voltage clamp and internal dialysis[13]. The internal dialysis fluid contained (in mM): 75.5 K-aspartate, 29 K_2HPO_4, 17.5 KH_2PO_4, 5.5 glucose (*p*H 7.0, *p*Ca 6.4). In some experiments (Fig. 2*a*, *b*), potassium salts were replaced by caesium salts. Except where otherwise noted, the external solution contained (in mM): 137 NaCl, 11.9 $NaHCO_3$, 0.42 NaH_2PO_4, 5.4 KCl, 1.05 $MgCl_2$, 1.0 $CaCl_2$, 5.5 glucose and 10 μM tetrodotoxin (TTX, *p*H 7.4). In these solutions, the resistance of the suction pipette tip was usually 0.8–1.2 MΩ, and the magnitude of the peak inward Ca current was a few nanoamperes. The resulting ohmic drop of a few millivolts was offset by series resistance compensation. Transmembrane voltage and current signals were recorded and analysed using a Digital Equipment 11/03 minicomputer. All experiments were done at room temperature (21 °C).

In all experiments, the membrane potential was held at −40 mV to inactivate any Na channels not blocked by 10 μM TTX. Figure 1 shows membrane currents associated with a series of 150-ms clamp pulses, spaced 10 mV apart. The magnitude of the peak inward current increases gradually with depolarization up to 0 mV as more and more Ca channels become activated (*a*). With pulses above +10 mV (Fig. 1*b*), the time-dependent inward current diminishes and eventually disappears at +70 mV. Beyond this apparent reversal potential, E_{rev}, the time-dependent current becomes increasingly outward. This pattern of current inversion was found consistently in each of the 70 cells we examined; with 1 mM Ca^{2+} outside and 150 mM K^+ inside, E_{rev} was always found near +60 or +70 mV.

Figure 1*b* illustrates the inversion of time-dependent current over the time scale of Ca channel inactivation. If the reversal is a property of current flow through Ca channels themselves, it should also be evident during the first few milliseconds following a clamp step, as the Ca channels open. Figure 1*c*, *d* shows records taken at a fast sampling rate in another single cell. The traces show membrane current after subtraction of linear capacity and 'leak' currents. Activation of inward Ca current becomes increasingly rapid over the range between −30 and +10 mV (*c*). This is in qualitative agreement with previous work on multicellular preparations (for example, ref. 8), but at all potentials the time-to-peak Ca current is briefer in these single cells. The family of records in panel *d* shows the behaviour at more positive potentials. The time-dependent current is maximally inward at +20 mV, virtually absent at +75 mV, and increasingly outward beyond this E_{rev}. Thus, the records are consistent with a true reversal of Ca channel current.

The ionic nature of the outward current at strongly positive potentials was explored by varying the ionic composition of the external or internal solutions. The outward current cannot be attributed to Ca efflux because intracellular free Ca^{2+} was kept low by the internal anions, and in some experiments (Fig. 2*a*), by internal EGTA as well. (The effectiveness of the Ca^{2+} buffering was indicated by the absence of contractile activity even at strong depolarizations). Furthermore, the decaying outward current cannot be explained by Na^+ efflux carried by electrogenic Na–Ca exchange as proposed by Mullins[9], since it was seen in the absence of internal or external Na^+ (Fig. 2*a*), or in the absence of external Ca^{2+} (Fig. 2*d*). The experiment illustrated by Fig. 2*a*, *b* went on to test the idea that the outward current is carried by K^+ ions. We found that the decaying outward current was abolished when Cs replaced K in the internal dialysate (*b*), as expected if K^+ ions carry current much more effectively than Cs^+ ions.

If potassium ions carry the outward current, does K^+ move through the Ca channel itself, or through another pathway such as a K-specific channel? One possibility is that K-efflux occurs through a Ca-activated K current, $I_{K(Ca)}$. This is a reasonable hypothesis since such currents have already been found in heart (for example, ref. 14), but it would require near-perfect cancellation of I_{Ca} and $I_{K(Ca)}$ at the apparent reversal potential (Fig. 1*d*), a result not found in systems where these currents are known to overlap (for example, ref. 15). Figure 2*c*–*e* provides direct evidence against an explanation involving $I_{K(Ca)}$. Panel *c* shows inward Ca current at 0 mV, and decaying outward current at +90 mV. Removal of extracellular Ca (*d*) abolishes the inward Ca current, and increases the background leak conductance, but leaves the decaying outward current at +90 mV virtually unchanged. This argues against $I_{K(Ca)}$ as the mechanism of the outward current, since removal of Ca_o should quench any intracellular Ca transient that occurs beneath the surface membrane despite the presence of internal Ca buffers. The ineffectiveness of external Ca removal contrasts with the immediate abolition of the time-dependent outward current by exposure to the inorganic calcium channel blocker Cd^{2+} (Fig. 2*e*; see also Fig. 4*d*–*f*).

0028–0836/82/230498—04$01.00

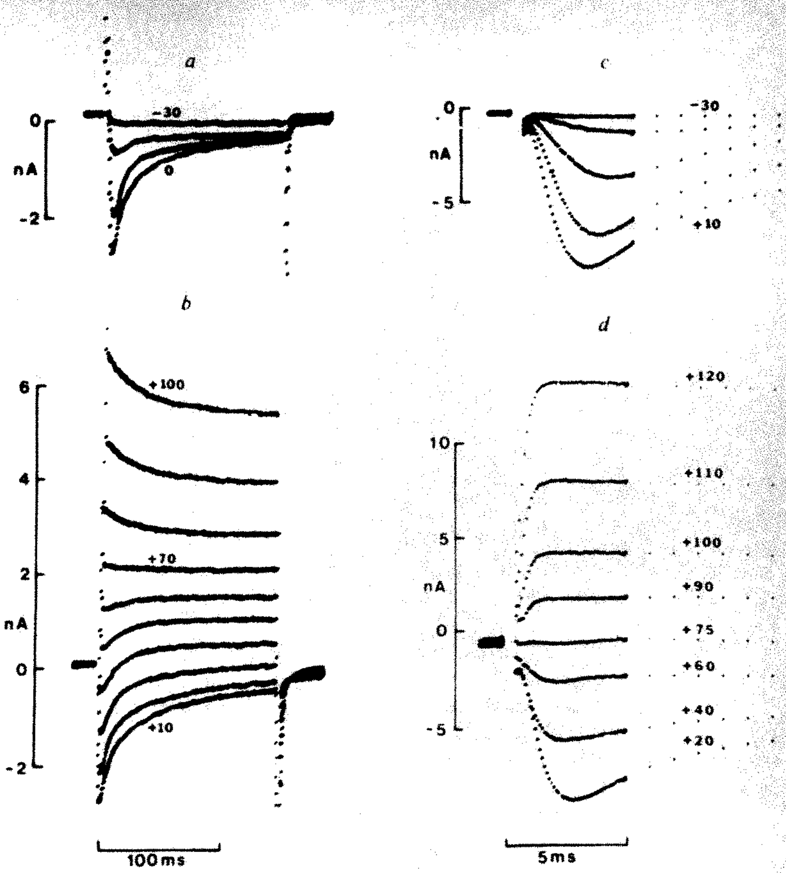


Fig. 1 Voltage and time-dependence of Ca channel current. *a*, *b*, Total membrane current records during a series of depolarizing pulses from a holding potential of −40 mV to levels spaced 10 mV apart. Pulses separated by at least 5 s. Capacity and leak current not subtracted. Current signal was filtered at 1 kHz and sampled at 1 kHz. Cell 64B. *c*, *d*, Membrane current records on a faster time scale from another cell, after digital subtraction of capacity transient and leak current. Current signals were filtered at 3 kHz. Sample rate was 10 kHz initially, then 1 kHz starting 5 ms after onset of voltage pulse. Linear 'leak' conductance and capacity current were estimated as the current during a hyperpolarizing pulse from −40 to −50 mV at the end of the run. 'Leak' conductance was 0.035 μS. The capacitative transient settled with a final time constant of 235 μs and its area corresponded to a total cell capacitance of ~184 pF. Data points are not shown for the first 300 μs after the onset of the voltage pulse while the membrane potential settled and the total current signal came into the range of the analog-to-digital converter. Cell 79G.

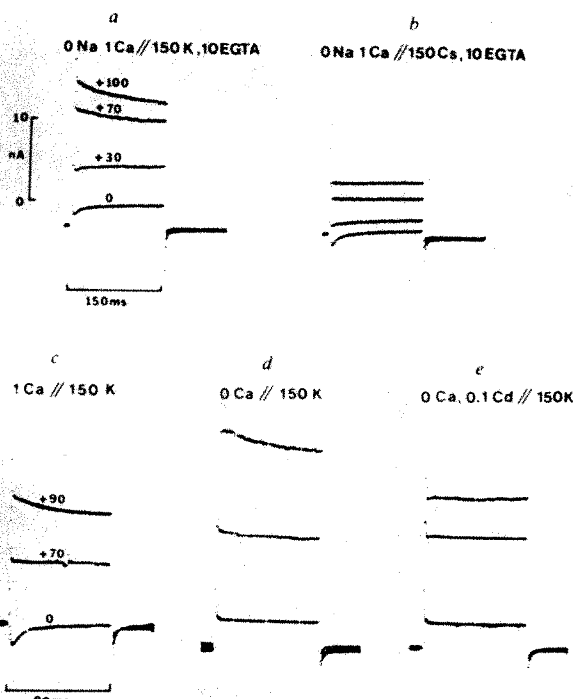


Fig. 2 Effects of removing internal K^+ or external Ca^{2+}. *a*, *b*: Effect of Cs^+ substitution for K^+ on decaying outward current. Total membrane currents recorded in 0-Na 0-K solution containing (in mM): 1 $CaCl_2$, 2.8 $MgCl_2$, 84 Cs-aspartate, 146 sucrose and 5.5 glucose, buffered to *p*H 7.4 with Tris. All internal solutions contain 10 mM K-EGTA. Addition of EGTA reduced the free Ca^{2+} concentration in the dialysate (kindly measured by Dr C. O. Lee using Ca^{2+}-sensitive microelectrodes) from ~0.4 μM to an undetectably low level. Voltage steps from a holding potential of −40 mV to levels indicated. *b*, Records taken with 150 Cs^+ in place of K^+ in internal solution; *a*, records taken 2 min after changing internal solution to the usual 150 K^+ dialysate. Cell 66B. *c–e*, Effects of changes in external divalent cation concentration from another experiment using solutions decribed in text: *c*, with 1 mM Ca outside; *d*, 0–Ca outside; *e*, 0–Ca + 100 μM Cd outside. Results in *d* taken after 90 s in 0–Ca solution. Cell 47B.

Can the decaying outward current be explained by the behaviour of a conventional voltage-dependent K channel? We found no reduction of the decaying outward current with 20 mM internal tetraethylammonium (TEA) or 1 mM external 4-aminopyridine (4-AP), levels that are usually effective at inhibiting K channels in heart and other excitable cells[16,17]. Unfortunately, higher doses of these agents were precluded by apparently nonspecific side-effects; 5 mM 4-AP produced a large outward holding current and increased leak conductance; even at 20 mM, intracellular TEA quickly diminished the inward Ca current and hastened the disappearance of all time-dependent current changes.

The most critical tests of whether the decaying outward current is carried by Ca channels came from experiments where Ca channels were blocked pharmacologically or inactivated by membrane depolarization. The idea behind the experiments was simple: if there is a genuine reversal potential where Ca channel current is zero because Ca influx and K efflux cancel, interventions that selectively inhibit Ca channels should have no effect at that potential. Figure 3 shows the effect of D600, a potent methoxy derivative of verapamil that is widely used as an inhibitor of Ca channels. A run before exposure to the drug shows a typical inversion of time-dependent current (*a*). After exposure to 0.3 μM D600 (*b*), both decaying inward and decaying outward currents were inhibited; the residual current in D600 is nearly linear and time-independent over the voltage range illustrated. Figure 3*c* superimposes current traces taken at three key test levels before and after D600. At the putative reversal potential, where time-dependent current vanishes, control and D600 traces coincide, just as expected if the driving force for net Ca channel current were zero and if D600 had no effect on leakage current.

The voltage-dependence of peak current and late current as obtained by this D600 dissection are plotted in Fig. 3*d*. Some features of the current–voltage relationships are surprising. First, late D600-sensitive current is appreciable, even though the predominant time-dependent current change seems largely complete at the end of the 150-ms clamp pulse (*c*). Second, persistent D600-sensitive current is significant even over the range between −50 and −30 mV, where there is very little

time-dependent inward current. Third, peak and late D600-sensitive currents grow quite large at strongly positive potentials; both I–V relationships show upward curvature between +30 and +100 mV.

The results illustrated in Fig. 3 are representative of a total of five experiments with 0.1–5 μM D600. The simplest interpretation is that Ca channel current truly reverses and that D600 blocks current flow in either direction. It could be argued that the null effect at E_{rev} is fortuitous, and that D600 not only blocks the Ca channel but also a voltage-dependent K channel. However, this explanation is very unlikely in view of the more extensive evidence illustrated in Fig. 4. Using the approach introduced in Fig. 3, we looked at varying degrees of Ca channel inhibition, not only by D600 (Fig. 4*a–c*), but also by the inorganic blocker Cd^{2+} (Fig. 4*d–f*), and by inactivation with depolarizing prepulses (Fig. 4*g–i*). In all three types of experiments, partial reduction of inward current below E_{rev} was associated with a partial decrease of outward current above E_{rev}, while the nearly time-independent current near E_{rev} remained essentially unchanged. These results provide strong evidence for a genuine reversal of Ca channel current at strong depolarizations. At the nearly physiological concentrations of Ca^{2+} and K^+ used in these experiments, the outward current is due to K^+ efflux through the channel, as originally hypothesized[8]. K^+ ions are not the only monovalent cations which can permeate the Ca channel. An outward movement of Cs^+ ions can also be detected in external media lacking permeant divalent cations or containing barium instead of calcium. Like the outward K^+ movement, the outward Cs^+ current can be inhibited by D600, Cd, or depolarizing prepulses; these interventions have no effect at the reversal potential seen with Ba^{2+} outside and Cs^+ inside, but reduce inward and outward currents recorded on either side of E_{rev}[18].

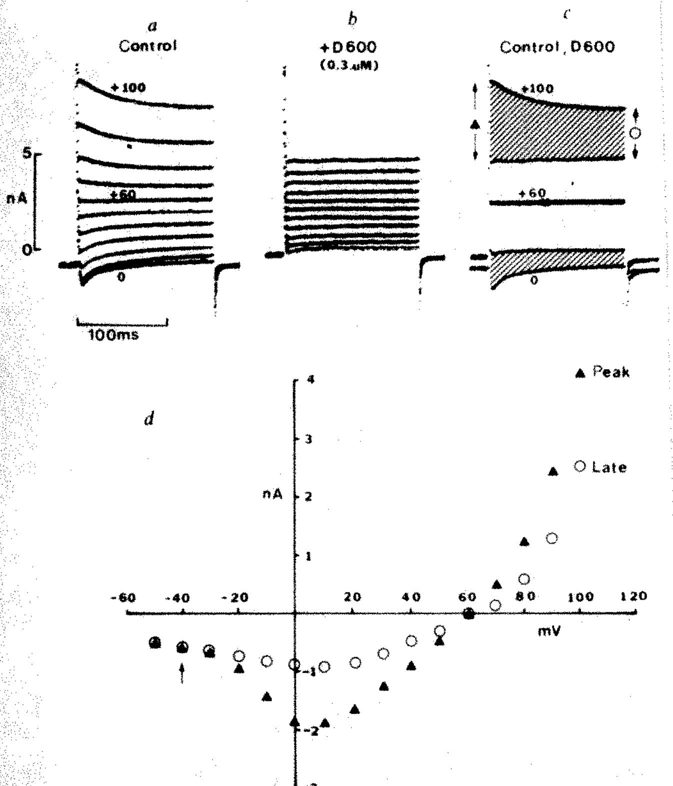


Fig. 3 Pharmacological dissection of calcium channel current using D600. *a*, Total membrane currents recorded during depolarizing pulses from −40 mV to levels spaced 10 mV apart, in conditions similar to Fig. 1*b*. *b*, Current records associated with the same set of voltage pulses after 5 min exposure to 0.3 μM D600. *c*, Repeat of current traces in *a* and *b*, for pulses to 0, +60 and +100 mV. The shaded areas indicate D600-sensitive current. Peak (▲) and late (○) D600-sensitive current at 100 mV were measured as indicated. *d*, Voltage-dependence of peak and late D600-sensitive current measured from signals shown in *a* and *b* and from records during pulses to levels below 0 mV. Arrow marks holding potential. Symbols at −40 and −50 mV denote changes in steady current. Cell 64A.

There is no conflict between these results in heart cells and earlier studies in dialysed molluscan nerve cells that did not show outward current through Ca channels[5–7]. The neurone experiments were done with Cs^+ or $Tris^+$ as the intracellular cations, with Ca^{2+} outside. In similar ionic conditions, our heart

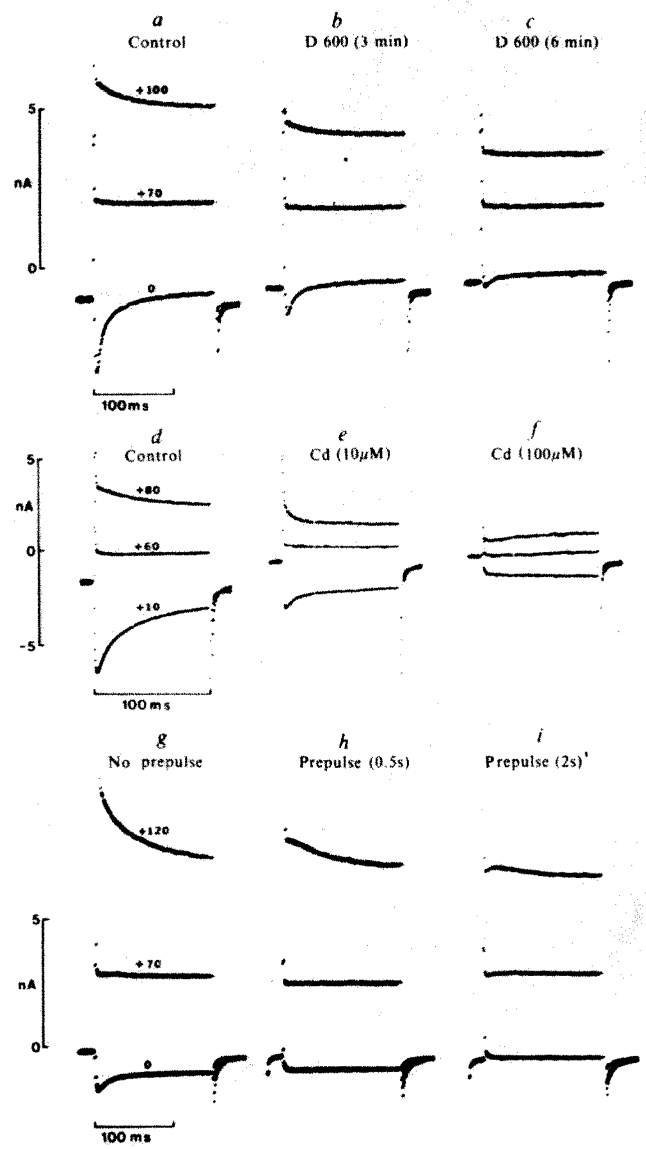


Fig. 4 Graded effects of interventions that reduce Ca channel conductance. *a–c*, Progressive development of D600 inhibition: *a*, 1 min after beginning D600 exposure, before any significant Ca channel block; *b*, after 3 min, when inward Ca current was roughly halved; *c*, after 6 min, when D600 effect had reached steady state. Cell 65C. Panels *d–f*, graded effect of external Cd^{2+}. The records in *d* and *e* were taken within a few minutes after increasing the external Cd^{2+} concentration, after block had reached steady state. In a total of six experiments using 10–500 μM Cd^{2+}, four cells showed no significant change in the level of current at the apparent reversal, one cell showed an upward shift, and one cell gave a downward shift. The Cd^{2+}-sensitive current resembled the D600-sensitive current illustrated in Fig. 3*d*. The speeding of inactivation evident in panel *e* was a consistent finding at low levels of Cd^{2+}. Cell 46A. Panels *g–i*, graded inactivation by conditioning prepulses to −10 mV. The current records were evoked by test pulses from a holding potential of −40 mV to the levels indicated. *g*, No prepulse preceding test pulses; *h*, 0.5 s conditioning prepulse, separated from test pulses by 20 ms repolarization to −40 mV; *i*, 2 s conditioning prepulse. Time-dependent current were reduced to a greater extent in *i* than in *h*; the Ca current recorded during prepulses to −10 mV showed a corresponding slow decline between 0.5 and 2 s. Cell 73A.

cells also lack decaying outward currents over the voltage range between 0 and +120 mV (Fig. 2*b*; ref. 18). It is interesting that outward Cs current can only be detected in the absence of external Ca. Presumably, Ca^{2+} ions bind to the channel much more strongly than Ba^{2+} ions[1], preventing significant outward movement of Cs^+, but not the efflux of the more permeant cation K^+.

These experiments show for the first time the feasibility of studying cardiac Ca channels while changing intracellular ions at will. As in the case of various patch clamp techniques[19,20], the suction pipette method offers its own unique combination of advantages. The relatively large pipettes used here give direct recordings of ensemble properties of ionic channels, like whole cell recordings using fine-tipped pipettes[19,21], but they also allow manipulation of the medium on the cytoplasmic side of the membrane, like cell-free patches[19,20]. Lability of Ca channel function does not seem to be a serious problem: inward and outward Ca channel currents often remain relatively stable for up to 30–40 min with external Ca and for even longer with external Ba. Systematic variations in intracellular constituents might give further improvements. But our experience already augurs well for future studies on modulation of Ca channels by adrenaline, acetylcholine or digitalis, and the role of putative intracellular mediators such as cyclic nucleotides or Ca^{2+} ions[2,3,22]. Intracellular Ca is also thought to promote Ca channel inactivation in a variety of excitable cells[23] including certain heart cells (for example, ref. 24). However, this cannot be the only mechanism since, as Fig. 2*d* shows, outward current through these Ca channels undergoes inactivation even in the absence of permeant divalent cations.

We thank Esther W. Lee for technical assistance and B. P. Bean and R. W. Aldrich for helpful discussion. This research was supported by the American Heart Association and the USPHS.

Received 18 January; accepted 14 April 1982.

1. Hagiwara, S. & Byerly, L. *A. Rev. Neurosci.* **4,** 69–125 (1981).
2. Tsien, R. W. & Siegelbaum, S. in *Physiology of Membrane Disorders* (eds Andreoli, T. E., Hoffman, J. F. & Fanestil, D. D.) 517–538 (Plenum, New York 1978).
3. Reuter, H. *A. Rev. Physiol.* **41,** 413–424 (1979).
4. Hille, B. in *Membranes: A Series of Advances* Vol. 3 (ed. Eisenman, G.) 255–323 (Dekker, New York, 1975).
5. Kostyuk, P. G. & Krishtal, O. A. *J. Physiol., Lond.* **270,** 545–568 (1977).
6. Akaike, N., Lee, K. S. & Brown, A. M. *J. gen. Physiol.* **71,** 509–531 (1978).
7. Hagiwara, S. in *Molluscan Nerve Cells: From Biophysics to Behavior* (eds Koester, J. & Byrne, J. H.) 33–54 (Cold Spring Harbor Laboratory, New York, 1980).
8. Reuter, H. & Scholz, H. *J. Physiol., Lond.* **264,** 17–47 (1977).
9. Mullins, L. *Am. J. Physiol.* **236,** C103–C110 (1979).
10. Lee, K. S., Kao, R. L. & Brown, A. M. *Circulation* **60,** 11–108 (1979).
11. Isenberg, G. & Klockner, U. *Nature* **284,** 358–360 (1980).
12. Powell, T., Terrar, D. A. & Twist, V. W. *J. Physiol., Lond.* **319,** 82–83P (1981).
13. Lee, K. S., Weeks, T. A., Kao, R. L., Akaike, N. & Brown, A. M. *Nature* **278,** 268–271 (1979).
14. Siegelbaum, S. A. & Tsien, R. W. *J. Physiol., Lond.* **299,** 485–506 (1980).
15. Meech, R. W. & Standen, N. B. *J. Physiol., Lond.* **249,** 211–239 (1975).
16. Kenyon, J. L. & Gibbons, W. R. *J. gen. Physiol.* **73,** 139–157 (1979).
17. Thompson, S. H. & Aldrich, R. W. in *The Cell Surface and Neuronal Function* (eds Cotman, C. W., Poste, G. & Nicolson, G. L.) 49–85 (North Holland, New York, 1980).
18. Lee, K. S., Lee, E. W. & Tsien, R. W. *Biophys. J.* **33,** 143a (1981).
19. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch. ges. Physiol.* **391,** 85–100 (1981).
20. Horn, R. & Patlak, J. B. *Proc. natn. Acad. Sci. U.S.A.* **77,** 6930–6934 (1980).
21. Hume, J. R. & Giles, W. *J. gen. Physiol.* (in the press).
22. Marban, E. & Tsien, R. W. *J. Physiol., Lond.* (in the press).
23. Eckert, R., Tillotson, D. & Brehm, P. *Fedn Proc.* **40** (in the press).
24. Marban, E. & Tsien, R. W. *Biophys. J.* **33,** 143a (1981).

Properties of single calcium channels in cardiac cell culture

H. Reuter

Department of Pharmacology, University of Berne, CH-3010 Berne, Switzerland

C. F. Stevens, R. W. Tsien & G. Yellen

Department of Physiology, Yale University, School of Medicine, New Haven, Connecticut 06510, USA

The permeability of cell membranes to Ca^{2+} ions is of great importance for a variety of cellular functions such as secretion of neurotransmitters and hormones, or contractile activation of heart cells. In several cell types including cardiac muscle, Ca^{2+} channels are gated by membrane potential and regulated by neurotransmitters[1–3]. We have used the patch clamp technique[4] in order to obtain information about the properties of single Ca channels in cultured heart cells. With Ba^{2+} carrying the inward current the single channel slope conductance is about 25 pS, and the probability of channel opening increases markedly with voltage. The channels become inactivated over a membrane potential range of about 60 mV. The mean channel open time is about 1 ms at 25 °C. However, a second much slower clustering behaviour primarily determines the slow activation time course of averaged currents. Isoprenaline does not affect single channel conductance, but seems to lengthen the mean open time and to decrease intervals between bursts.

Tissue cultured myocytes were obtained from neonatal rat hearts[5,6]. After seeding, the cells were incubated for 2–4 days in a culture medium containing 0.1 mM 8-bromo cyclic AMP to increase the likelihood of detecting single Ca channels[7]. Most cells used for our experiments had spindle-like shapes and showed striations. Single-channel currents were recorded from cell-attached membrane patches with seals between 10 and 100 GΩ (ref. 4). The recording pipette contained 96 mM $BaCl_2$ solution with 20 μM tetrodotoxin (buffered at *p*H 7.4 with 10 mM HEPES); thus, the current carrier was Ba^{2+} rather than Ca^{2+}. The composition of the bathing solution was (mM): NaCl 137, KCl 5.4, $MgCl_2$ 1, $CaCl_2$ 0.02, glucose 10, HEPES buffer 10 (*p*H 7.4); temperature 24–26 °C. Consistent results have been obtained from experiments on 24 patches with Ca channel activity, most of which had only a single Ca channel present. Our analysis of single Ca channel properties presented here is based primarily on results obtained from five patches, each of which provided rather complete information; three of these had only a single Ca channel present.

Calcium channels switch randomly between open and closed states, and appear not to dwell at intermediate conductance levels; some depolarizations produce one or more channel openings of various durations, whereas for others, the channel fails to open at all (Fig. 1).

The steady-state probability that a channel will be open depends on membrane potential, as does the current that flows while the channel is open (Fig. 2). We have estimated the single channel current (i) by eye (Fig. 2*a*) and calculated open probability (p) from

$$I = Npi \qquad (1)$$

where the mean current I was found by time-averaging the current flowing through calcium channels over 50 or 100 ms, and N is the number of channels functioning in the patch ($N = 1$ for Fig. 2). The probability of a channel being open increases with depolarization, at a maximum rate of about e-fold per 10 mV, from nearly zero at a membrane potential of −30 mV to 0.7 at around +30 mV. Over this same voltage range, the open channel current–voltage relationship is approximately ohmic with a slope conductance of about 25 pS (isotonic $BaCl_2$ extracellularly).

During long depolarizing steps Ca channels show inactivation that is 20 or 30 times slower than activation (Fig. 3*a*). We studied steady-state inactivation by measuring Ca channel currents at a fixed potential after holding the membrane potential at various conditioning levels (Fig. 3*b*). This inactivation occurs through a decrease in the probability that a channel is available to open, and not through a decrease in single channel current, *i*. Steady-state inactivation increases with conditioning depolarization and can be fitted by a Boltzmann distribution. Channels are half-inactivated at a membrane potential of about

−35 mV, while channels just begin to be activated with depolarizations slightly greater than this (about −30 mV).

We have estimated the average duration of a channel opening, $\bar{t}$, by the equation

$$\bar{t} = \frac{p}{r} \qquad (2)$$

where p is the fraction of the time a channel spends open (from equation (1)) and r is the average opening rate obtained by counting the number of opening transitions n during the period T over which the time-averaged current was calculated ($r = n/T$). When necessary the average opening rate was corrected for openings too brief to be detected; this correction assumed an exponential distribution of open times, and was at most 1%. The mean open time was always close to 1 ms. For example, for the experiment that yielded the records shown in Fig. 2,

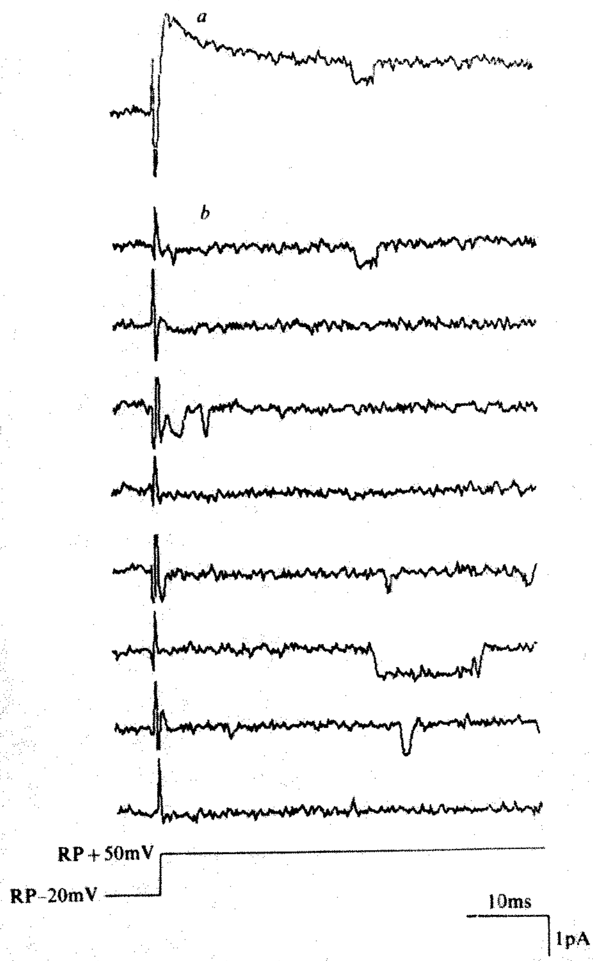


Fig. 1 Single Ca channel currents from cultured rat heart cells. In record *a*, the capacitive transient was reduced by analog compensation[4]. The eight sequential records in *b* have had the remaining capacitive transient and the leakage current subtracted (the first record in *b* is derived from *a*). Records were inspected visually, and portions containing channel activity or obvious artefacts were excluded from the leakage current average. Membrane potential of the patch was held at 20 mV negative to the resting potential (RP) and stepped to RP+50 mV. Resting potential of this cell was estimated to be −50 mV; this was done by comparing unit currents through Ca-activated inward current channels[5] in these cell-attached patches with the unit current carried by Ba^{2+} ions through such channels in detached patches, for which the true membrane potential is known (D. Colquhoun, E. Neher, H.R. and C.F.S., unpublished results). These channels were frequently observed in these experiments, but in the voltage range where we have studied Ca channel current they carry only a small unit current; thus, they can easily be distinguished from the Ca channel currents and make no appreciable contribution to the observed current.

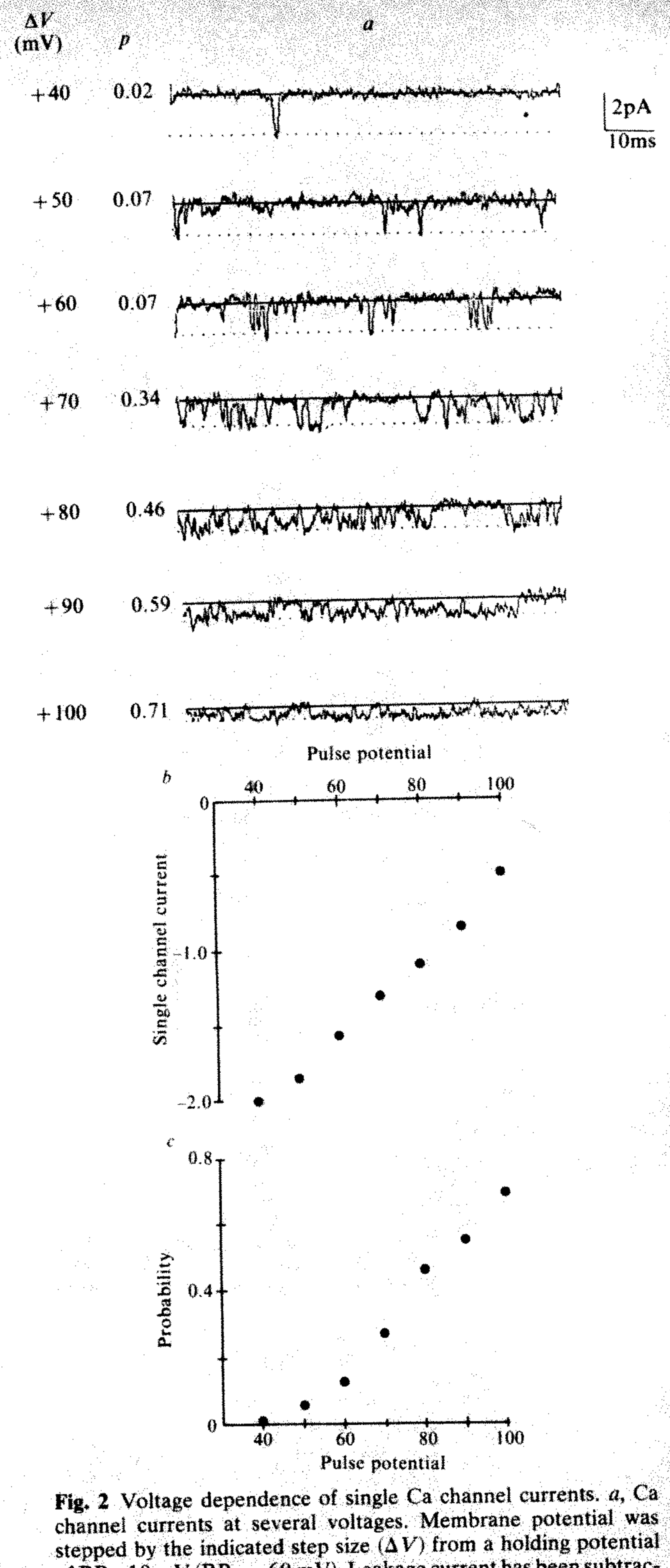


Fig. 2 Voltage dependence of single Ca channel currents. *a*, Ca channel currents at several voltages. Membrane potential was stepped by the indicated step size (ΔV) from a holding potential of RP − 10 mV (RP ≈ −60 mV). Leakage current has been subtracted from these records. The solid lines indicate the baseline after leak subtraction. The dotted lines indicate the measured open-channel current. Beside each record is the average probability (p) that the channel was open, as determined from the individual record shown (see text, equation (1)). This patch had only a single active Ca channel. Data were filtered at 1 kHz. *b*, Voltage-dependence of open channel current. Open channel current was fitted visually for the experiment in *a* for each membrane potential. A pulse potential of 40 mV corresponds approximately to a membrane potential of −30 mV, as above. *c*, Voltage dependence of Ca channel gating. The time-averaged probability that the single active Ca channel in this patch was open was determined (for each voltage) by dividing the average current by the open channel current (equation (1)). Average current was measured over a 90 ms interval beginning 5 ms after the start of each pulse.

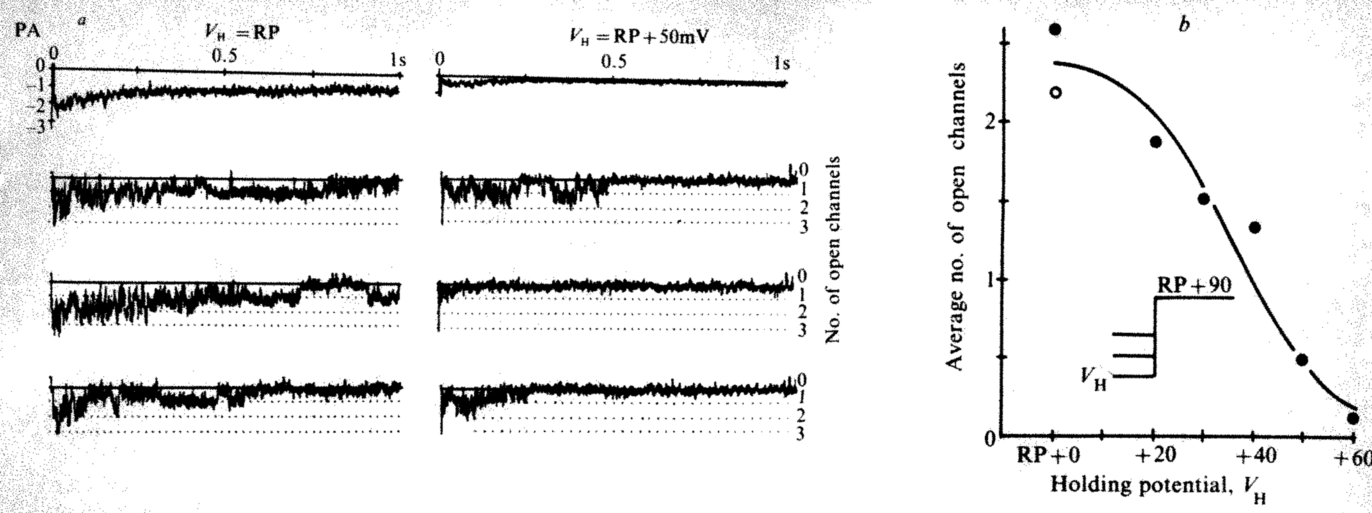


Fig. 3 Inactivation of Ca channel currents. *a*, Ca channel currents at a fixed potential, stepping from two different holding potentials. Left, V_H (holding potential) = RP; right, V_H = RP + 50 mV (RP ≈ −70 mV). Command potential is RP + 90 mV in both cases. At top is the average current, calculated from about 20 individual records, for each holding potential (leakage current subtracted). There is no significant change in time-to-peak when the averaged records are scaled to the same amplitude. Below are examples of three individual records from each holding potential (same scale as averages). Dotted lines on the individual records are spaced 0.86 pA apart, which is the size of a single channel current at RP + 90 mV. Notice the additional clustering which occurs on a very slow (hundreds of milliseconds) time scale. This patch showed a maximum of four open channels for large depolarizations. Data were filtered at 1 kHz. *b*, Voltage dependence of Ca channel inactivation. The time-averaged number of open channels in the first 50 ms of a test pulse to PR + 90 mV is plotted as a function of holding potential. The average number of open channels is found by dividing the average current during the first 50 ms of the test pulse by the unit open channel current which did not vary with holding potential. The solid line is a Boltzmann distribution:

$$N = N_0[1 + \exp((V_m - V_{0.5})/k)]^{-1}$$

with parameters chosen for the best visual fit: N_0 = 2.4 open channels; $V_{0.5}$ = RP + 35 mV; k = 10 mV. The open circle is a repeat of the V_H = RP point, done at the end of the inactivation series. Observations of Ca channel activation before and after the inactivation experiment illustrated here agree closely with the activation curve in Fig. 2*b*.

the mean open time increased from 0.65 ms at a pulse potential of 60 mV to 1.4 ms for an 80 mV pulse. Although the standard errors of these measurements are less than 10% of the mean, we estimate that systematic errors of closer to 15% are present due to uncertainties in counting opening transitions (the interested reader should count openings in the records of Fig. 2). Our values for mean open time are therefore close to the estimates reported for Ca channels in snail neurones[8] and bovine chromaffin cells[9].

The Ca channels studied here exhibit distinctive gating behaviour. Their openings are not dispersed at random throughout the period of depolarization, but rather tend to occur in groups of two or more closely spaced openings separated by longer silent times (see Fig. 2*a*, +60 and +70 mV). It is as if the channel enters an 'activated' condition in which it rapidly flickers open and closed. The average current is activated (see Fig. 3*a*, upper row) with a time constant on the order of 10 ms, and this longer time seems to reflect the burst length rather than the characteristic time (~1 ms) of the individual openings.

The β-adrenergic agonist isoprenaline, mimicking the action of the hormone adrenaline, causes an increase in cardiac calcium currents[1,2,10]. We have investigated the effects of this drug at the single channel level in several preliminary experiments. Isoprenaline caused no increase in the single channel current, but in one experiment (10 μM) it produced a 60% increase in the mean open time which was manifest as a corresponding increase in the probability from 0.06 to 0.09 ($V_m \approx +10$ mV). A similar but smaller effect was observed in a second experiment (1 μM). If this effect is confirmed by a more complete analysis, isoprenaline can be viewed as causing an increased conductance[10], at least partly, by altering the fraction of time an 'activated' channel spends in its open state (see ref. 2).

The kinetic properties of single Ca channel currents qualitatively agree with previously reported kinetic data on global Ca currents in a variety of cardiac tissues[1]. We find a steep increase in the probability of channels opening over the same potential range where activation of Ca conductance occurs[1,11]. Our single channel records exhibit steady-state inactivation kinetics similar to those reported for global Ca currents[1,11]. However, in contrast to previous results there is a smaller range of overlap between potentials where channels open and become inactivated. This may partly be due to the fact that we have used isotonic $BaCl_2$ solution in our pipette which is known to alter surface charges[12] and thereby shift activation and inactivation ranges along the voltage axis. We have used Ba^{2+} instead of Ca^{2+} ions because of their higher permeability through Ca channels[3,8].

Our measurements of single channel conductance yield values that are more than an order of magnitude larger than those estimated from fluctuation analysis in molluscan neurones in comparable conditions[13]. However, they are consistent with two reports on single Ca channels in snail neurones[8], and bovine chromaffin cells[9]. In the latter note, bursts of channel openings, presumably similar to those observed in our study, have also been mentioned.

In comparison with Na channels[14,15], Ca channels show a much more complex kinetic behaviour. There is a fast gating process (mean channel open time ~1 ms at 25 °C) and a slower one that results in bursts. It is interesting to speculate that one of the gating processes is regulated by some intermediate metabolic step, such as phosphorylation of the Ca channel. In this respect it will be important to pursue our preliminary isoprenaline results, since β-adrenergic catecholamines are thought to regulate the Ca conductance through such a mechanism[1,2].

This work was performed in the Department of Pharmacology, University of Berne and was supported by the Roche Research Foundation and the Swiss NSF. We thank C. Becker for providing tissue culture cells, Dr A. Cachelin for help with some of the experiments, and Roche Animal Farm, Füllinsdorf for the gift of pregnant rats.

Received 21 December 1981; accepted 31 March 1982.

1. Reuter, H. *A. Rev. Physiol.* **41,** 413–424 (1979).
2. Tsien, R. W. & Siegelbaum, S. in *Physiology of Membrane Disorders* (eds Andreoli, T. E., Hoffman, J. F. & Fanestil, D. D.) 517–538 (Plenum, New York, 1978).
3. Hagiwara, S. & Byerly, L. *A. Rev. Neurosci.* **4,** 69–125 (1981).
4. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, J. F. *Pflügers Arch. ges. Physiol.* **391,** 85–100 (1981).
5. Colquhoun, D., Neher, E., Reuter, H. & Stevens, C. F. *Nature* **294,** 752–754 (1981).
6. Mark, G. E. & Strasser, F. F. *Expl Cell Res.* **44,** 217–233 (1966).

7. Rotter, A., Ray, R. & Nirenberg, M. Fedn Proc. 38, 476a (1979).
8. Lux, H. D. & Nagy, K. Pflügers Arch. ges. Physiol. 391, 252–254 (1981).
9. Fenwick, E. M., Marty, A. & Neher, E. J. Physiol., Lond. 319, 100P–101P (1981).
10. Reuter, H. & Scholz, H. J. Physiol., Lond. 264, 49–62 (1977).
11. Reuter, H. & Scholz, H. J. Physiol., Lond. 264, 17–47 (1977).
12. Hille, B., Woodhull, A. M. & Shapiro, B. I. Phil. Trans. R. Soc. B270, 301–318 (1975).
13. Krishtal, O. A., Pidoplichko, V. I. & Shakhovalov, Y. A. J. Physiol., Lond. 310, 423–434 (1981).
14. Sigworth, F. J. & Neher, E. Nature 287, 447–449 (1980).
15. Horn, R., Patlak, J. & Stevens, C. F. Nature 291, 426–427 (1981).

Calcium-binding protein parvalbumin is associated with fast contracting muscle fibres

M. R. Celio & C. W. Heizmann

Anatomisches Institut der Universität Zürich, Gloriastrasse 19, CH-8006 Zürich, and Institut für Pharmakologie und Biochemie der Universität Zürich, Winterthurerstrasse 260, CH-8057 Zürich, Switzerland

The basic mechanism underlying the contraction–relaxation cycle of vertebrate muscles is a Ca^{2+} exchange between the sarcoplasmic reticulum and the myofibrils. Relaxation is achieved by retrieval of Ca^{2+} from the myofibrils and transfer to the sarcoplasmic reticulum. However, it is uncertain whether the rate of Ca^{2+} uptake by the sarcoplasmic reticulum can account for the known speed of relaxation[1,2]. In fast contracting muscles the Ca^{2+}-binding protein parvalbumin is postulated to facilitate relaxation[3–8]. Using immunohistochemical techniques, we show here that parvalbumin is located exclusively in type II (fast-twitch) mammalian skeletal muscle fibres which can be further subdivided into five subgroups displaying distinct staining intensities. As the active state decays more rapidly in fast than in slow muscles[9,10], our results support the contention that parvalbumin may be concerned with rapid muscle relaxation[3–8] and in addition suggest a range of relaxation properties in muscle fibres belonging to the same histochemical fibre type.

Newborn and adult Wistar rats of both sexes were killed and various skeletal muscles quickly dissected out, stretched and frozen in isopentane cooled by liquid nitrogen. The muscles were then divided and the respective halves separately processed for histochemical and immunohistochemical localization. Additionally, heart muscle tissue taken from the atria and ventricles, smooth muscle tissue of the gut, great vessels and non-pregnant uteri, as well as specialized, small cross-striated muscles with a high intrinsic speed of contraction (external eye muscles and intrinsic laryngeal muscles[9]) were dissected out, fixed in Bouin fluid and embedded in paraffin. The high solubility of parvalbumin prevented the use of unfixed or postfixed cryostat sections for the immunohistochemistry. Monospecific antisera to rat muscle parvalbumin were used in dilutions up to 1:30,000 in phosphate-buffered saline (PBS). The specificity of the antisera has been described elsewhere[11]; there is no cross-reaction[12] with troponin-C, calmodulin or S-100 protein. Sections were further processed according to the unlabelled antibody technique of Sternberger[13]. Briefly, goat-anti-rabbit IgG (Nordic) was applied for 30 min, followed by peroxidase–antiperoxidase complex (Sternberger–Meyer Inc.). The peroxidase-dependent reaction was detected by means of diaminobenzidine HCl/H_2O_2. Control slides were processed identically and simultaneously with preimmune sera of the same animal and with preadsorbed antisera.

Parvalbumin immunoreactivity in the extensor digitorum longus muscle of newborn rats appears on postnatal days 3 or 4. In adult rats, anti-parvalbumin antisera stain extensor digitorum longus extrafusal muscle fibres in at least five different degrees of intensity (Fig. 1*a*). In the soleus muscle the parvalbumin immunostaining is much less conspicuous (Fig. 1*b*). The parvalbumin immunostaining pattern (Fig. 2*a*) was correlated with histochemical fibre typing by staining serial sections for myofibrillar ATPase[14,15] after preincubation at *p*H 4.3 (Fig. 2*b*) and with formalin followed by *p*H 10.6 (Fig. 2*c*). Several laboratories have shown near-equivalence of formalin-based and acid-based subdivisions of fast fibres[16–18]. Accordingly, formaldehyde-resistant fast fibres will here be designated type IIA, and formaldehyde-sensitive ones type IIB. The muscle fibres could thus be histochemically classified into the main groups: type I (Fig. 2*b*) (slow-twitch, oxidative), type IIA (fast-twitch, oxidative-glycolytic) and type IIB (fast-twitch, glycolytic)[14,15,19] (Fig. 2*c*). The type IIB fibres showed the strongest immunoreactivity towards parvalbumin antisera (Fig. 2*a*) and could be differentiated into two subgroups showing graded

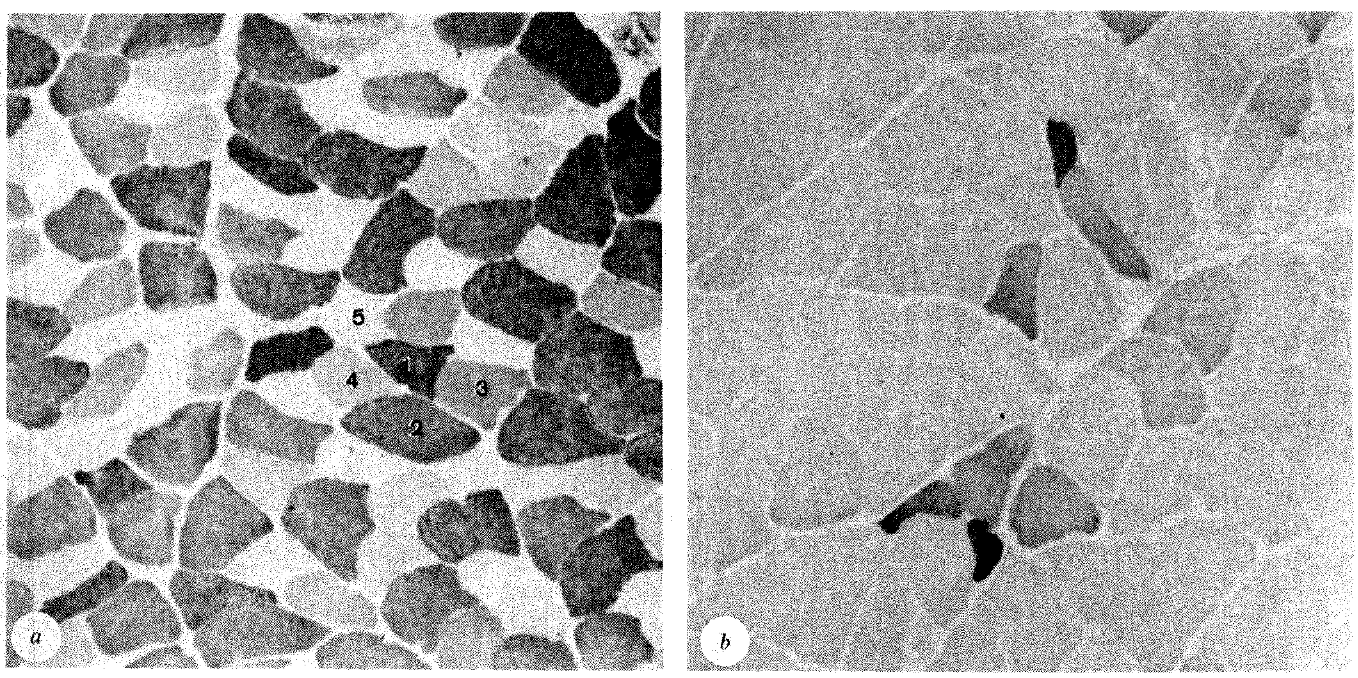


Fig. 1 *a*, Survey of a transverse section through the belly of the musculus extensor digitorum longus (EDL; fast contracting/fast relaxing[9]) of an adult rat incubated with parvalbumin antiserum and processed by the peroxidase–antiperoxidase technique[13]. The parvalbumin immunostaining shows a chequered appearance with at least five different shadowings in staining intensity, even at the extreme antiserum dilution of 1:20,000. *b* Shows a cross-section of the ipsilateral soleus (slow contracting/slow relaxing[9]) of the same animal processed identically. In the soleus very few muscle fibres are labelled by the parvalbumin antiserum. ×150.

0028–0836/82/230504—03$01.00

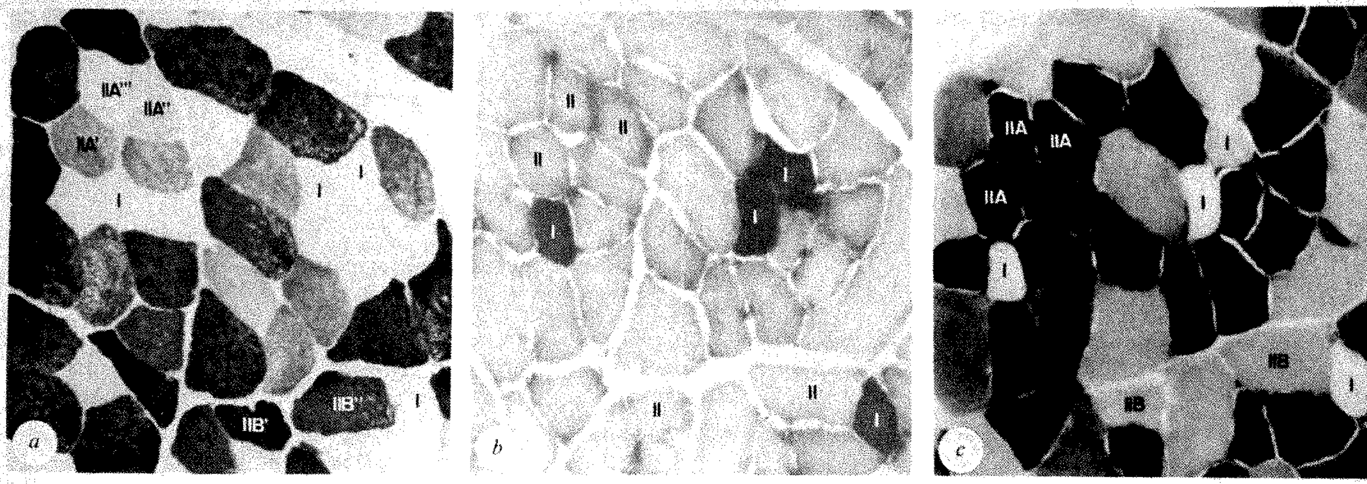


Fig. 2 Consecutive transverse sections of the EDL of an adult rat, analysed for parvalbumin immunoreactivity (*a*) and enzyme-histochemical fibre types (*b*, *c*). Adult Wistar rats (Ivanovas) were killed by a blow to the head and the muscles quickly dissected out, frozen in isopentane cooled by liquid nitrogen and halved with a precooled razor blade. One part was freeze-substituted with acetone for 1 week at −70 °C and fixed at the same temperature with a mixture of acetone (90 ml), formalin (40% per 5 ml), acetic acid (1 ml) and water (4 ml) for another week. The specimens were thawed, embedded in Paraplast and the sections incubated with antisera against rat muscle parvalbumin in dilutions varying between 1:5,000 and 1:30,000 for 48 h in a moist chamber at 4 °C. The immune complex was visualized by the unlabelled antibody technique of Sternberger[13]. Identical distribution and localization of parvalbumin immunoreactivity were achieved after perfusion fixation of the rat's body with Bouin fluid and embedding of the muscles in paraffin (see Fig. 4 legend). The second half of the muscle was cut in a cryostat and the sections processed for acid-stable (*p*H 4.3) ATPase activity (*b*) and formalin/alkali-stable (*p*H 10.6) ATPase activity (*c*), following established procedures[14–18]. Corresponding areas were photographed and muscle fibres identified. Type IIB fibres (*b*, *c*) display two distinct degrees of staining intensity with parvalbumin antisera (IIB', IIB'' in *a*) while type IIA fibres (*b*, *c*) may even show three (IIA', IIA'', IIA''' in *a*). IIA''' fibres are actually devoid of parvalbumin immunoreactivity and in this respect resemble type I fibres. Note that the strongest immunolabelling is detectable in the small-calibre IIB' fibres (*a*). ×148.

differences in staining intensity. The type IIA muscle fibres were moderately stained in three distinct degrees of intensity (Fig. 2*a*). Type I muscle fibres always remained nonreactive (Fig. 2*a*). In muscles spindles, one nuclear chain fibre was immunolabelled, whereas bag fibres were not (Fig. 3). High parvalbumin immunoreactivity was detectable in fibres belonging to the global portion of the musculus rectus lateralis oculi (Fig. 4*b*) and in the musculus thyreoarytenoideus lateralis (Fig. 4*a*). The slow tonic fibres in the orbital layer of external eye muscles (Fig. 4*b*) as well as the musculus thyreoarytenoideus medialis (vocalis) were nonreactive (Fig. 4*a*). Cardiac muscle fibres and smooth muscle cells were never labelled by parvalbumin antisera (not shown).

The immunostaining was reproducible for a wide range of antiserum dilutions. Identical correlations with histochemical fibre typing were obtained using other fixatives and embedding procedures (see figure legends). The staining pattern was homogeneous and evenly distributed over the whole muscle fibre in longitudinal section (not shown) and in cross-section.

A discussion of the role of parvalbumin in muscle hinges on the hypothesis that parvalbumin facilitates the relaxation by shuttling Ca^{2+} from troponin-C to the sarcoplasmic reticulum[3–8]. The sarcoplasmic reticulum is actually well developed in type II muscle fibres[20], which are rich in parvalbumin, and type II muscle fibres form the bulk of fast muscles, which as a whole are indeed fast relaxing. Our observation of a high parvalbumin content in such fibres supports the notion that parvalbumin is implicated in muscle relaxation, as the decay of the active state is shorter in fast than slow motor units. This assumption is strengthened by the observation that parvalbumin immunoreactivity appears on days 3 and 4 after birth in the extensor digitorum longus, coincident with changes in contraction properties, in particular the decrease in the relaxation time after a twitch or tetanus[21,22]. Similarly, it is probable that the parvalbumin-positive intrafusal nuclear chain fibre relaxes faster than do nuclear bag fibres[23].

The teleological meaning of rapid relaxation is that muscle fibres are sooner ready for a new contraction. It is therefore not surprising that parvalbumin immunoreactivity is particularly high in fibres of muscles characterized by very high contraction frequencies. In the rat[9,10] these muscles are the external eye muscles and the inner laryngeal muscles, both displaying the highest intrinsic speed of shortening and the shortest twitch half-relaxation times.

Type I muscle fibres show a relatively poorly-developed sarcoplasmic reticulum and lack parvalbumin immunoreactivity. Like cardiac muscle, they are well provided with mitochondria which may be involved in Ca^{2+} homeostasis[24]. As no isoforms of parvalbumin have been found in rat skeletal muscles[25], the relative staining intensities presumably correspond to differences in the concentration of parvalbumin within these muscle fibres and may reflect a continuum in relaxation properties. This is consistent with the finding of metabolic subpopulations of type II fibres and the proposition of a dynamic continuum of muscle fibres[26–31] correlated with the wide spectrum of contraction times[32,33].

An incidental effect attributed to parvalbumin is that it might, by releasing heat as it takes up Ca^{2+}, contribute substantially to the 'unexplained' heat output displayed by many muscles early in tetanus[34]. This hypothesis may now be tested by comparing muscles with high and low parvalbumin concentrations.

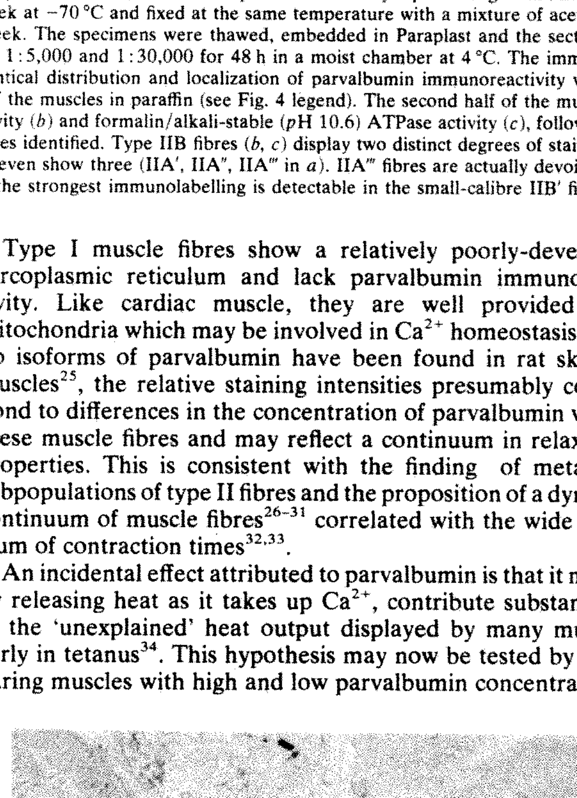


Fig. 3 Cross-section through the intracapsular polar region of a muscle spindle of the soleus muscle. Four miniature muscle fibres are suspended in the spindle capsule. One chain fibre (c) is strongly parvalbumin-reactive whereas the second chain fibre (c) as well as the bag fibres (b) are only weakly labelled. The parvalbumin immunostaining of soleus muscle spindles was always unequivocal, in contrast to that of identically treated EDL muscle spindles. I and II: types I and II extrafusal muscle fibres (as determined in following sections treated for the histochemical detection of ATPase). ×250.

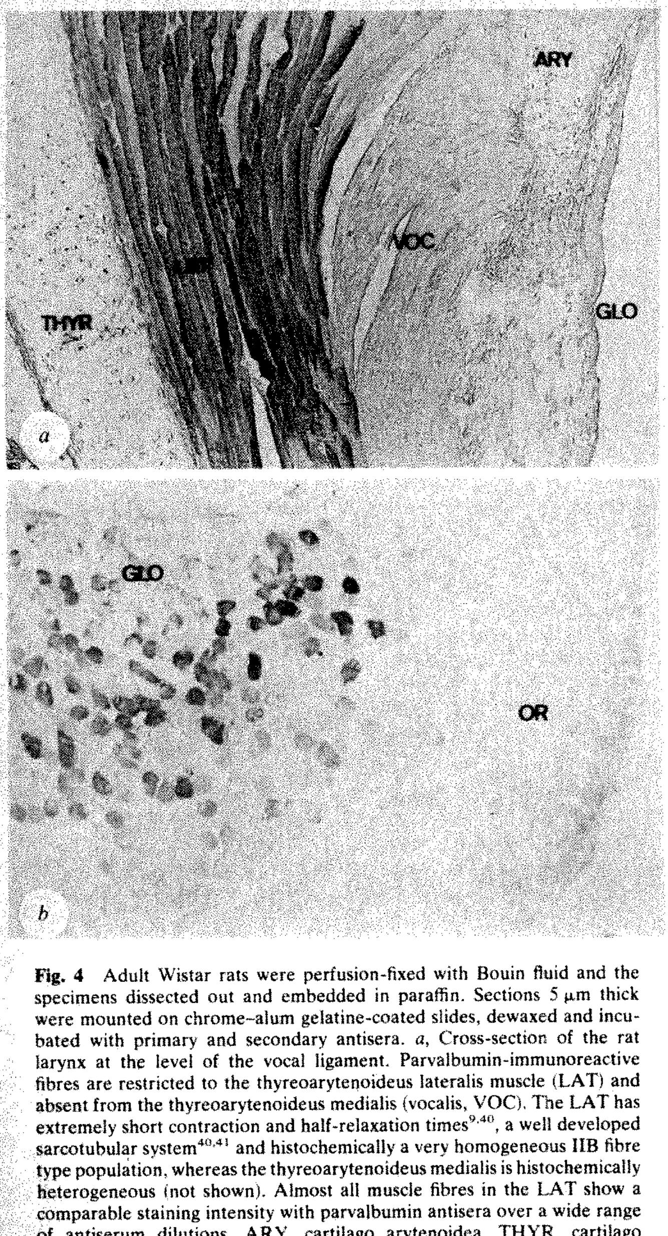


Fig. 4 Adult Wistar rats were perfusion-fixed with Bouin fluid and the specimens dissected out and embedded in paraffin. Sections 5 μm thick were mounted on chrome-alum gelatine-coated slides, dewaxed and incubated with primary and secondary antisera. *a*, Cross-section of the rat larynx at the level of the vocal ligament. Parvalbumin-immunoreactive fibres are restricted to the thyreoarytenoideus lateralis muscle (LAT) and absent from the thyreoarytenoideus medialis (vocalis, VOC). The LAT has extremely short contraction and half-relaxation times[9,40], a well developed sarcotubular system[40,41] and histochemically a very homogeneous IIB fibre type population, whereas the thyreoarytenoideus medialis is histochemically heterogeneous (not shown). Almost all muscle fibres in the LAT show a comparable staining intensity with parvalbumin antisera over a wide range of antiserum dilutions. ARY, cartilago arytenoidea, THYR, cartilago thyreoidea, GLO, glottis. ×125. *b*, Cross-section of an external eye muscle. Only fibres restricted to the global portion (GLO) of the rectus lateralis muscle are recognized by parvalbumin antisera, whereas the small-calibre fibres of the orbital layer (OR), with a known slower contraction time[42], remain unlabelled. The intensity of immunostaining with parvalbumin antisera varies among muscle fibres, reflecting the presence of a wide range of different histochemical fibre types[43]. ×100.

Our results demonstrate an association of the Ca^{2+}-binding protein parvalbumin with the fast-contracting muscle fibres of rat muscles, in agreement with earlier findings on fish muscles[35–39]. Parvalbumin could therefore represent a reliable system for Ca^{2+} homeostasis, allowing rapid cyclic relaxation in phasically active muscle fibres. Furthermore, parvalbumin immunohistology seems to be a very sensitive auxiliary technique to enzyme histochemistry and myosin immunohistology for classifying muscle fibres into various subtypes.

We thank Drs E. Jenny and W. Zenker for supporting this work, Drs A. Rowlerson and M. Müntener for valuable discussion and Mrs H. Tedaldi, Mrs H. Weber and Miss Chr. Hemmerle for technical assistance.

Received 18 January; accepted 31 March 1982.

1. Weber, A., Herz, R. & Reiss, I. *Proc. R. Soc.* **B160,** 489–499 (1964).
2. Ebashi, S. *A. Rev. Physiol.* **38,** 293–313 (1976).
3. Gerday, C. & Gillis, J. -M. *J. Physiol., Lond.* **258,** 96P–97P (1976).
4. Gillis, J.-M. & Gerday, C. in *Calcium Binding Proteins and Calcium Function* (eds Wasserman, R. H. *et al.*) 193–196 (Elsevier, Amsterdam, 1977).
5. Fischer, E. H. *et al.* in *Molecular Basis of Motility* (eds Heilmeyer, L. M. G. Jr, Ruegg, J.-C. & Wieland, T.) 137–153 (Springer, Berlin, 1976).
6. Pechère, J. F., Derancourt, J. & Haiech, J. *FEBS Lett.* **75,** 111–114 (1977).
7. Blum, H. E., Lehky, P., Kohler, L., Stein, E. A. & Fischer, E. H. *J. biol. Chem.* **252,** 2834–2838 (1977).
8. Gillis, J.-M. in *Calcium Binding Proteins, Structure and Function* (eds Siegel, F. L., Carafoli, E., Kretsinger, R. H., MacLennan, D. M. & Wasserman, R. H.) 309–311 (Elsevier, Amsterdam, 1980).
9. Close, R. I. *Physiol. Rev.* **52,** 129–197 (1972).
10. Luff, A. R. *J. Physiol., Lond.* **313,** 161–171 (1981).
11. Celio, M. R. & Heizmann, C. W. *Nature* **293,** 300–302 (1981).
12. Van Edlik, L. J. & Watterson, D. M. *J. biol. Chem.* **256,** 4205–4210 (1981).
13. Sternberger, L. A. in *Immunocytochemistry*, 104–169 (Wiley , New York, 1979).
14. Guth, L. & Samaha, F. T. *Expl Neurol.* **28,** 365–367 (1970).
15. Brooke, M. H. & Kaiser, K. K. *Archs Neurol.* **23,** 369–379 (1970).
16. Tunnel, G. L. & Hart, M. N. *Archs Neurol.* **34,** 171–173 (1977).
17. Edjtehadi, G. & Lewis, D. M. *J. Physiol., Lond.* **287,** 439–453 (1979).
18. Spurway, N. C. *J. Histochem.* **29,** 87–88 (1981).
19. Peter, J. B., Barnard, R. J., Edgerton, V. R., Gillespie, C. A. & Stempel, K. E. *Biochemistry* **11,** 2627–2633 (1972).
20. Gauthier, G. F. *Z. Zellforsch.* **95,** 462–482 (1969).
21. Drachman, D. B. & Johnston, D. M. *J. Physiol., Lond.* **234,** 29–42 (1973).
22. Close, R. *J. Physiol., Lond.* **173,** 74–95 (1964).
23. Boyd, I. A. *Prog. Brain Res.* **44,** 33–49 (1976).
24. Carafoli, E. in *Calcium in Cell Regulation* (ed. Smellie, R. S.) 101–103 (Biochemistry Society, London, 1974).
25. Berchtold, M. W., Wilson, K. J. & Heizmann, C. W. (in preparation).
26. Spamer, C. & Pette, D. *Histochemistry* **60,** 9–19 (1979).
27. Nemeth, P. M., Hofer, H. W. & Pette, D. *Histochemistry* **63,** 191–201 (1979).
28. Kugelberg, E. *J. neurol. Sci.* **27,** 269–289 (1976).
29. Jansson, E., Sjödin, B. & Tesch, P. *Acta physiol. scand.* **104,** 235–237 (1978).
30. Billeter, R., Heizmann, C. W., Howald, H. & Jenny, E. *Eur. J. Biochem.* **116,** 389–395 (1981).
31. Billeter, R. *et al. Histochemistry* **65,** 249–259 (1980).
32. Bárány, M. *J. gen. Physiol.* **50,** 197–216 (1967).
33. Garnett, R. A. F., O'Donovan, M. J., Stephens, J. A. & Taylor, A. J. *J. Physiol., Lond.* **287,** 33–43 (1978).
34. Homsher, E. & Irving, M. *Adv. physiol. Sci.* **5,** 141–150 (1980).
35. Pechère, J. F., Capony, J. P. & Demaille, J. *Syst. Zool.* **22,** 533–548 (1973).
36. Hamoir, G. *Symp. Biol. Hung.* **17,** 17–33 (1974).
37. Gosselin-Rey, C. in *Calcium Binding Proteins* (eds Drabikowski, W., Strzelecka-Golaszewska, H. & Carafoli, E.) 643–677 (Elsevier, Amsterdam, 1974).
38. Gerday, C., Joris, B., Gerardin-Otthiers, N., Collin, S. & Hamoir, G. *Biochimie* **61,** 589–599 (1979).
39. Hamoir, G. & Gerardin-Otthiers, N. *Comp. Biochem. Physiol.* **65B,** 199–206 (1980).
40. Takahashi, S. *Comp. Biochem. Physiol.* **62A,** 539–544 (1979).
41. Schmalbruch, H. *Z. Zellforsch.* **119,** 120–146 (1971).
42. Bach-Y-Rita, P. & Ito, F. *J. gen Physiol.* **49,** 1177–1198 (1966).
43. Vita, G. F., Mastaglia, F. L. & Johnson, M. A. *Neuropath. appl. Neurobiol.* **6,** 449–463 (1980).

Genetic regulation of resistance to intracellular pathogens

Emil Skamene*, Philippe Gros*, Adrien Forget†, Patricia A. L. Kongshavn‡, Carole St Charles* & Benjamin A Taylor§

* Montreal General Hospital Research Institute, Montreal, Quebec H3G 1A4, Canada
† Département de Microbiologie et Immunologie, Université de Montréal, Montreal, Quebec H3C 3J7, Canada
‡ Department of Physiology, McGill University, Montreal, Quebec H3G 1Y6, Canada
§ Jackson Laboratory, Bar Harbor, Maine 04609, USA

Natural resistance of mice to infections with *Salmonella typhimurium* and *Leishmania donovani* is regulated by chromosome 1 gene(s) designated *Ity* and *Lsh*, respectively[1,2]. Given the fact that these two microorganisms are taxonomically and antigenically distinct, and yet the host response to them is regulated by the same locus or complex[3,4], one might expect that the resistance to other intracellular pathogens would be controlled similarly. Innate resistance of inbred mice to infection with *Mycobacterium bovis* (BCG) is regulated by a single, dominant, autosomal gene designated *Bcg* which is known to exist in two allelic forms: resistant Bcg^r and susceptible Bcg^s (ref. 5). The distribution of Bcg^r and Bcg^s alleles, among a total of 14 inbred and 38 recombinant inbred (BXD and BXH) strains, matches exactly that established for resistant and susceptible alleles of *Lsh* (gene controlling resistance to *Leishmania donovani*) and *Ity* (gene controlling resistance to *Salmonella typhimurium*), suggesting that resistance to all these pathogens is controlled by the same chromosome 1 locus. The existence of such a chromosomal locus is further supported by the *Bcg–Lsh* and *Bcg–Ity* linkage as established by formal backcross analysis.

0028–0836/82/230506—04$01.00

Mice of selected inbred and recombinant inbred strains were infected intravenously with a dispersed inoculum of BCG[6] at a dose of $2-3 \times 10^4$ colony-forming units (CFU) of BCG. The number of viable BCG organisms in the spleen was determined 3 weeks after infection. All the inbred mouse strains which were tested segregated into two distinct groups with respect to the host response to BCG infection: resistant strains (*Bcg*^{r}), in which practically no proliferation of BCG inoculum occurred, and susceptible strains (*Bcg*^{s}), which allowed up to 100-fold BCG multiplication in their spleens (Table 1). Segregation into resistant and susceptible strains matches exactly that which was established for *Ity*- and *Lsh*-regulated resistance to *S. typhimurium* and *L. donovani*[7,8].

Recombinant inbred strains BXD and BXH (originating from the BCG-susceptible C57BL/6J progenitor, and the BCG-resistant DBA/2J and C3H/HeJ progenitors, respectively), when infected with BCG, also segregated into two distinct groups (Figs 1, 2). The strain distribution pattern (SDP) of *Bcg*^{r} and *Bcg*^{s} alleles among BXD and BXH strains was found to be completely concordant with that established for *Ity*^{r}, *Ity*^{s} and *Lsh*^{r}, *Lsh*^{s} alleles[3,4]. One strain deserves a comment: BXH-2 typed as susceptible to BCG while reported to be *Lsh*^{r} and *Ity*^{r}. However, the (C57BL/6J × BXH-2)F_1 hybrid mice were found to be resistant to BCG ($\log_{10}$ BCG $= 4.69 \pm 0.14$), thus clearly indicating that the phenotypic susceptibility to BCG of BXH-2 strain is caused by a factor unrelated to the *Bcg* gene that overrides the resistance conferred by the *Bcg*^{r} allele. It is of interest that the BXH-2 strain was also found to express high levels of murine leukaemia virus, and has a high incidence of lymphoma. The basis for these characteristics is not fully understood, but was acquired after the strain was already inbred[9].

We have also examined the linkage between *Lsh* and *Bcg* in a backcross population. F_1 hybrids between B10.A (*Bcg*^{s} *Lsh*^{s}) and A/J (*Bcg*^{r} *Lsh*^{r}) were mated to the susceptible B10.A parent to produce backcross progeny. These were infected with BCG, and splenectomized 3 weeks later for *Bcg* typing. The animals were then infected with *L. donovani* after a further 3 weeks and the *Lsh* type was established in their livers[8]. Preliminary experiments confirmed that neither splenectomy nor previous infection with BCG influenced the *Lsh* typing. These manipulations did not change the susceptibility to *L. donovani* of B10.A control mice, nor did they change the resistance to *L. donovani* to (B.10.A × A)F_1 control mice. Of 26 [(B10.A × A)F_1) × B10.A] backcross animals, 12 were typed as resistant to BCG and 14 were typed as susceptible to BCG, consistent ($\chi^2 = 0.15$, $P > 0.7$) with the hypothesis that the trait of resistance to BCG is under monogenic, dominant control[5]. Furthermore, all 12 of the BCG-resistant backcross mice proved to be resistant to *L. donovani*, and all of the BCG-susceptible backcross mice proved to be susceptible to *L. donovani*. A similar experiment to establish the linkage of *Ity* with *Bcg* could not be done since splenectomy influenced the *Ity* typing: splenectomized mice of *Ity*^{r} strains became

Table 1 *Bcg*, *Ity* and *Lsh* typing of inbred strains of mice

Strain	*Bcg**		*Ity*†	*Lsh*‡
A/J	3.90±0.40	r	r	r
AKR/J	3.95±0.21	r	r	r
C3H/HeN	3.95±0.15	r	r	r
C3H/HeJ	3.71±0.11	r	r§	r
C57BR/J	4.12±0.25	r	r	r
C57L/J	4.34±0.22	r	r	r
CBA/J	4.21±0.13	r	r	r
DBA/2J	3.93±0.18	r	r	r
129/J	4.13±0.70	r	–	r
BALB/cJ	5.80±0.08	s	s	s
DBA/1J	6.18±0.13	s	s	s
C57BL/6J	5.75±0.09	s	s	s
C57BL/10J	6.05±0.20	s	s	s
CE/J	5.67±0.14	s	–	s

r, Resistant; s, susceptible. All mice except C3H/HeN were purchased from the Jackson Laboratory, Bar Harbor, Maine. C3H/HeN mice originating at The Charles River Breeding Laboratories, Wilmington, Massachusetts, were purchased from Canadian Breeders, St Constant, Quebec.

* $\log_{10}$ CFU of BCG±s.d. in spleens 3 weeks after infection with 10^4 CFU BCG.

† *Ity* typing taken from Robson and Vas[13], Plant and Glynn[7] and O'Brien *et al.*[3].

‡ *Lsh* typing taken from Bradley *et al.*[8].

§ C3H/HeJ is *Ity*^{r} (ref. 14) although phenotypically susceptible to *S. typhimurium* due to the presence of *Lps*^{d} mutation in that strain.

Table 2 Linkage of *Bcg* with *Lsh* in segregating backcross progeny

Population		*Bcg* $\log_{10}$ of BCG CFU	*Bcg* Type	*Lsh* Type	*Lsh* LDU
(B10.A × A)F_1	1	4.43	r	r	8
	2	3.69	r	r	0
	3	4.11	r	r	29
	4	3.97	r	r	0
	5	3.67	r	r	14
B10.A	1	5.91	s	s	1,782
	2	5.70	s	s	1,297
	3	5.62	s	s	1,396
	4	5.76	s	s	1,664
	5	5.56	s	s	644
F_1 × B10.A backcross	1	3.85	r	r	0
	2	4.06	r	r	16
	3	5.72	s	s	575
	4	4.26	r	r	6
	5	5.61	s	s	791
	6	3.88	r	r	26
	7	5.36	s	s	210
	8	5.90	s	–	Dead
	9	3.82	r	r	43
	10	4.03	r	r	10
	11	3.45	r	r	39
	12	5.32	s	s	640
	13	5.15	s	s	1,332
	14	4.14	r	r	8
	15	5.08	s	s	441
	16	5.92	s	s	955
	17	5.46	s	s	550
	18	4.32	r	r	15
	19	4.11	r	r	13
	20	5.68	s	s	548
	21	5.36	s	s	464
	22	6.13	s	s	580
	23	6.12	s	s	214
	24	4.08	r	r	6
	25	3.70	r	r	14
	26	6.45	s	s	587

r = Resistant; s = susceptible. *Bcg* typing performed on spleens 3 weeks after intravenous infection with 3×10^4 CFU BCG[5]. *Lsh* typing performed on livers 2 weeks after intravenous infection according to Bradley[8] and expressed as Leishman–Donovan units (LDU) per liver according to Actor[15]. An animal typed as *Bcg*^{r} (resistant) if the $\log_{10}$ BCG count in the spleen was <4.5 (2 s.d. above the mean BCG count of resistant (B10.A × A)F_1 controls, 95% confidence limit). An animal typed as *Lsh*^{r} (resistant) if the LDU in the liver was <46 (2 s.d. above the mean LDU count of resistant (B10.A × A)F_1 controls, 95% confidence limit). The LDU counts were generally somewhat lower and their range was greater in F_1 × B10.A backcross susceptible mice than in B10.A parents. Although the *Lsh* gene undoubtedly dominates the regulation of resistance to *L. donovani* these results suggest that other genes also have some effect in controlling the acute growth rate of this parasite. Same conclusion was already reached by Bradley[8].

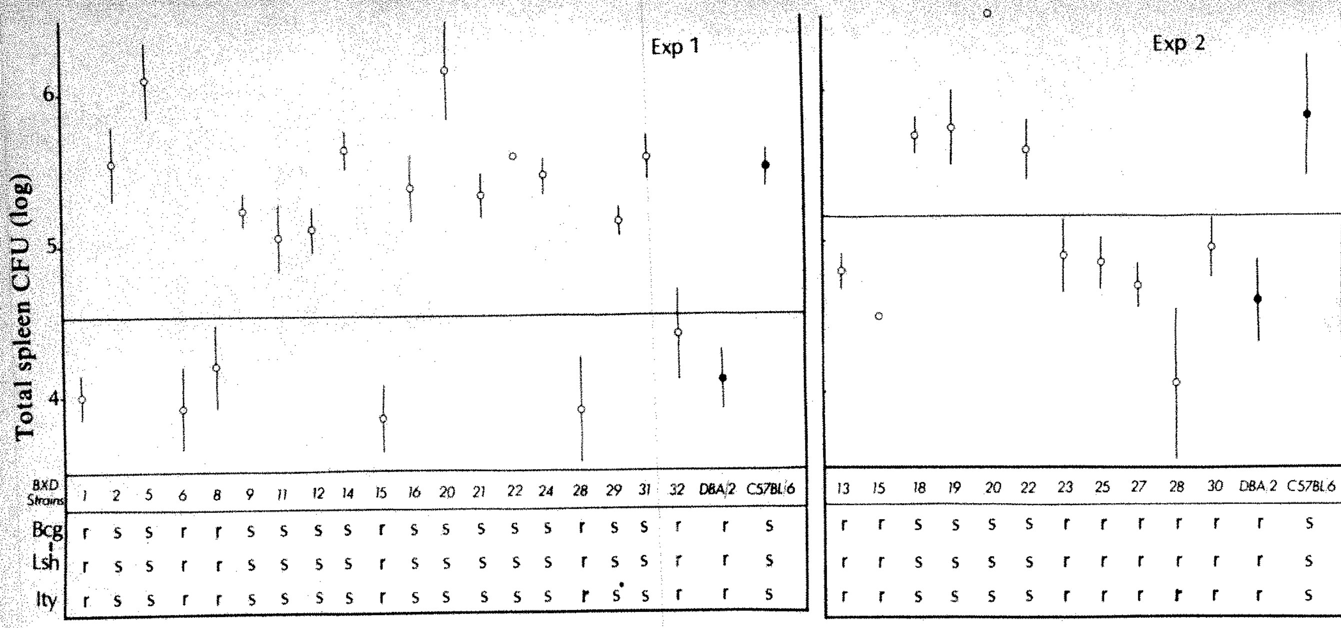


Fig. 1 Typing of BXD RI strains for resistance to BCG. Two to eight animals of each recombinant inbred strain and six animals of control DBA/2 and C57BL/6 strains were infected with 2–3 × 10^4 CFU BCG in two separate experiments. Geometric means ± s.d. of viable BCG in the spleens of infected animals were calculated 3 weeks after the infection. The horizontal line represents 95% confidence limit (two standard deviations above the mean of BCG numbers in the spleens of resistant DBA/2 progenitors) for typing of strains as *Bcg*^{r} (below the line) or *Bcg*^{s} (above the line). *Lsh* type of BXD strains was taken from Bradley *et al.*[2] except for BXD-28 and BXD-20. *Lsh*^{s} allele of these strains, identified by Plant *et al.*[4], was confirmed by us according to the criteria described in Table 2 (mean LDU 4,445 ± 57 got BXD-18 and 7,059 ± 1,066 for BXD-20, with values of LDU 116 ± 25 for *Lsh*^{r} control and 11,355 ± 2,455 for *Lsh*^{s} control in that particular experiment). *Ity* types of BXD strains were taken from O'Brien *et al.*[3] except for BXD-28 and BXD-29. Designation of *Ity* alleles of these two strains is taken from Plant *et al.*[4]. The *Ity*^{r} type of BXD-28 strains was confirmed according to the criteria described in Table 3, by typing (BXD-28 × B10.A)F_1 mice (7/7 animals survived the typing dose of 10^4 *Salmonella* i.v.). r, Resistant; s, susceptible.

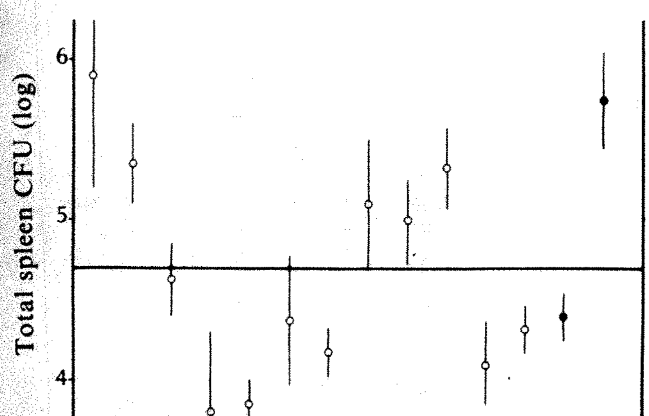


Fig. 2 Typing of BXH RI strains for resistance to BCG. Details as for Fig. 1, except for the resistant progenitor which in this case is C3H/HeJ. r, Resistant; s, susceptible; ?, some of the BXH strains were not *Ity* typed because of the interfering influence of the *Lps*^{d} mutation in the genome of C3H/HeJ progenitor; *, *Ity* type established by A. D. O'Brien (personal communication). Note that BXH-2 is designated as *Bcg*^{r} although phenotypically susceptible for BCG (see explanation in the text). BXH-10 is designated as *Lsh*^{s} according to the criteria described in Table 2 (mean LDU 990 ± 290), in variance with Bradley *et al.*[2]. *Ity*^{s} allele of BXH-10 strain was established by typing (BXH-10 × B10.A)F_1 mice, according to the criteria described in Table 3 (0/8 animals survived the typing dose of 10^4 *Salmonella* i.v.).

phenotypically susceptible to *S. typhimurium* infections. The reverse experiment (*Ity* typing first, with subsequent *Bcg* typing) also proved to be technically impossible. *Ity–Bcg* linkage was, therefore, established by progeny testing which eliminated the necessity of splenectomy and of double infection of an individual animal. Twenty-five randomly chosen males of the [(B10.A × A)F_1 × B10.A)] backcross generation were mated to *Bcg*^{s}–*Ity*^{s} females of B10.A strain. The resulting progeny were *Ity*-typed and the corresponding males of the parental generation were subsequently *Bcg*-typed. We reasoned that, if the *Bcg* and *Ity* genes were identical or closely linked, the offspring of BCG-susceptible (*Bcg*^{s}/*Bcg*^{s}, homozygous) backcross males mated with *Bcg*^{s}*Ity*^{s} females should breed true and all should type as *Salmonella*-susceptible (*Ity*^{s}/*Ity*^{s}, homozygotes). The offspring of BCG-resistant (*Bcg*^{r}/*Bcg*^{s}, heterozygous) backcross males and B10.A females should segregate between *Salmonella*-susceptible (*Ity*^{s}/*Ity*^{s}, homozygotes) and *Salmonella*-resistant (*Ity*^{r}/*Ity*^{s}, heterozygotes) progeny.

Due to a laboratory accident, eight of the backcross males (BC 18–BC 25) were lost before they could be *Bcg*-typed. Fortunately, two litters of their mating with B10.A females were available and the *Bcg* type of the males could, therefore, be deduced from *Bcg* typing of the litter. When the results of *Bcg* and *Ity* typing were matched, a *Bcg*^{r}–*Ity*^{r} or *Bcg*^{s}–*Ity*^{s} concordance was found among 24 out of 25 first backcross males. The only discordance, BC 22, (*Bcg*^{s}(?)–*Ity*^{r}) is inconclusive because the *Bcg* typing was based on only six progeny, and could have resulted from chance segregation ($P = 0.016$), even if the BC 22 male had been *Bcg*^{r}/*Bcg*^{s}.

Collectively, the data presented here clearly establish that *Bcg*, the gene controlling resistance to *Mycobacterium bovis* BCG, maps on chromosome 1 and that it is either identical to, or linked with, the gene(s) controlling resistance to *S. typhimurium* and *L. donovani*. Since no conclusive recombination between the *Bcg*, *Ity* and *Lsh* alleles was found by back-

Table 3 Linkage of *Bcg* with *Ity* by progeny testing

	BCG infection*			*Salmonella* infection†	
	Total spleen BCG (log)	No. resistant/total B10.A × BC progeny§	*Bcg* type	*Ity* type	No. resistant/total B10.A × BC progeny§
BC 1‡	4.11		r	r	2/7
BC 2	5.72		s	s	0/14
BC 3	4.92		r	r	5/13
BC 4	5.62		s	s	0/3
BC 5	5.40		s	s	0/7
BC 6	3.94		r	r	3/7
BC 7	4.53		r	r	3/7
BC 8	3.94		r	r	5/10
BC 9	4.22		r	r	4/5
BC 10	5.33		s	s	0/8
BC 11	4.68		r	r	4/8
BC 12	5.54		s	s	0/8
BC 13	4.58		r	r	4/12
BC 14	4.54		r	r	9/12
BC 15	5.64		s	s	0/4
BC 16	4.36		r	r	6/10
BC 17	4.54		r	r	8/12
BC 18		0/7	s	s	0/8
BC 19		1/5	r	r	2/5
BC 20		1/4	r	r	5/8
BC 21		1/5	r	r	4/9
BC 22		0/6	s	r	4/11
BC 23		3/4	r	r	5/9
BC 24		0/6	s	s	0/9
BC 25		0/7	s	s	0/6
B10.A	5.60–5.92‖		s	s	0/25
(B10.A × A)F_1	3.93–4.64‖		r	r	25/25

* Animals typed 3 weeks after intravenous infection with 2.5×10^4 CFU BCG. Resistant (r): $\log_{10}$ BCG count in the spleen <mean ± 2 s.d. of resistant control, 95% confidence limit. Any animal with $\log_{10}$ BCG count above that value typed as susceptible (s).
† Animals typed 1 week after intravenous infection of 10^4 *Salmonella typhimurium* as r = resistant (survivors) or s = susceptible (dead).
‡ BC = [(B10.A × A)F_1 × B10.A].
§ B10.A × BC = B10.A × [(B10.A × A)F_1 × B10.A]. Two successive litters of the same parental pair were typed in some instances.
‖ Range of control values in 5 individual experiments which constitute the table.

cross analysis, and because of the full concordance of SDP of resistant and susceptible alleles of *Bcg*, *Ity* and *Lsh* among the total of 38 RI strains examined, these data suggest that the genetic control of resistance to *M. bovis* BCG, *S. typhimurium* and *L. donovani* is exerted by a single gene with pleiotropic effect. An alternative interpretation, that of a complex of closely-linked genes[3], seems less likely although it obviously cannot be ruled out completely. The fact that the susceptible allele is recessive, and that it is limited to a relatively few inbred strains, is more compatible with the possibility that the susceptibility gene might represent a mutation and be of a recent origin, rather than being a variant of a polymorphic gene that occurs in natural populations. That being the case, a single mutation from resistance to susceptibility would be more likely than the accumulation of three independent mutations to susceptibility genes (Bcg^s–Lsh^s–Ity^s) in coupling on the same chromosome.

The phenotypic expression of the gene in question is, so far, recognized *in vivo* only: in each case the resistant allele expresses itself as slow (or null) net growth rate of the particular microorganism in the liver or spleen[5,10,11] early in the course of infection, most probably before the specific immune response is established. It seems likely that the cellular expression of the resistant allele(s) is a bacteriostatic effect of the intracellular milieu of the parasitized macrophage[12], although no clear *in vitro* data are available as yet. The normal (resistant) allele might have other functions (either in macrophages or other cell types) but it would not be surprising if its chief function were the defence against intracellular parasites. Accordingly, the discovery of the mechanism by which this gene exerts its effect could reasonably be expected to shed considerable light on how macrophages help to control infections by organisms of medical significance. A consideration of common features (for example, biochemical or nutritional requirements for multiplication) of these three taxonomically distinct pathogens (*L. donovani*, *S. typhimurium* and *M. bovis*) could provide a starting point in this endeavour. We propose that natural resistance to other infections, such as tuberculosis and leprosy, to name but two, is also regulated by this gene. The biological implications of these findings for identification of genetically susceptible individuals and for the development of appropriate strategies for their protection and treatment are obvious.

This research was supported, in part by NIH contract N01-CB-84269, MRC grants 5389, 5448 and 6431 and grant GM18684 from the National Institute of General Medical Sciences. The Jackson Laboratory is fully accredited by the American Association for Accreditation of Laboratory Animal Care.

Received 15 December 1981; accepted 31 March 1982.

1. Plant, J. & Glynn, A. A. *Clin. exp. Immun.* **37,** 1–6 (1979).
2. Bradley, D. J., Taylor, B. A., Blackwell, J., Evans, E. P. & Freeman, J. *Clin. exp. Immun.* **37,** 7–14 (1979).
3. O'Brien, A. D., Rosenstreich, D. L. & Taylor, B. A. *Nature* **287,** 440–442 (1980).
4. Plant, J. E., Blackwell, J. M., O'Brien, A. D., Bradley, D. J. & Glynn, A. A. *Nature* **297,** 510–511 (1982).
5. Gros, P., Skamene, E. & Forget, A. J. *Immunology* **127,** 2417–2421 (1981).
6. Forget, A., Skamene, E., Gros, P., Miailhe, A. C. & Turcotte, R. *Infect. Immun.* **32,** 42–47 (1981).
7. Plant, J. & Glynn, A. A. *J. infect. Dis.* **133,** 72–78 (1976).
8. Bradley, D. J. *Clin. exp. Immun.* **30,** 130–140 (1977).
9. Bedigian, H. G., Taylor, B. A. & Meier, H. *J. Virol.* **39,** 632–640 (1981).
10. Hormaeche, C. E. *Immunology* **37,** 311–318 (1979).
11. Bradley, D. J. *Acta tropica* **36,** 171–179 (1972).
12. Hormaeche, C. E. *Immunology* **41,** 973–979 (1980).
13. Robson, H. G. & Vas, S. J. *J. infect. Dis.* **126,** 378–386 (1972).
14. O'Brien, A. D., Rosenstreich, D. L., Metcalf, E. S. & Scher, I. in *Genetic Control of Natural Resistance to Infection and Malignancy* (eds Skamene E., Kongshavn, P. A. L. & Landy, M.) 101 (Academic, New York, 1980).
15. Actor, P. *Expl Parasit.* **10,** 1–9 (1960).

Are the *Lsh* and *Ity* disease resistance genes at one locus on mouse chromosome 1?

Janet E. Plant*, Jenefer M. Blackwell†, Alison D. O'Brien‡, D. J. Bradley† & A. A. Glynn*

* Central Public Health Laboratory, Colindale Avenue, London NW9 5HT, UK
† Ross Institute of Tropical Hygiene, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK
‡ Department of Microbiology, Uniformed Services University of the Health Services, 4301 Jones Bridge Road, Bethesda, Maryland 20014, USA

***Ity*, a gene controlling natural resistance to *Salmonella typhimurium* infection[1,2], and *Lsh*, which controls innate resistance to *Leishmania donovani* infection[3], are both located on chromosome 1 of the mouse[4,5] and it has been suggested[6] that they might be identical. O'Brien *et al.*[7] examined salmonella resistance in recombinant inbred (RI) strains of mice, and observed discordant responses in three strains (BXD-18, BXD-20, BXD-29) from one progenitor strain combination (C57BL/6J × DBA/2J) when compared with the published *Lsh* typing[5]. On the basis of this and additional information on salmonella resistance in (C57BL/6J × BXD-18)F$_1$ and [(C57BL/6J × DBA/2J) × BXD-29] backcross mice, it was concluded that *Ity* and *Lsh* are closely linked but distinct genetic loci. We have now re-examined both leishmania and salmonella resistance in larger numbers of these putative recombinant strains and in various hybrid and backcross generations, and our results indicate that BXD-18 and BXD-20 retype as *Lsh*^{s} and BXD-29 as *Ity*^{s}. Hence there is no longer any discordance between *Lsh* and *Ity* typings for the RI strains. The results do, however, suggest additional modifying genetic control of salmonella resistance in BXD strains involving at least two other genetic loci. There is therefore no clear evidence to demonstrate that *Ity* and *Lsh* are not the same genetic locus.**

Initially we examined leishmania and salmonella resistance in the three BXD strains thought to differ at *Lsh* and *Ity* and in an additional strain (BXD-28) found earlier[7] to be intermediate in its susceptibility to *S. typhimurium* TML strain. (The BXD mice used in our experiments were derived from stock sent by Dr Benjamin Taylor of the Jackson Laboratory.) Salmonella resistance was further examined in the progeny of BXD-29 mice mated with stock DBA/2, C57BL and (C57BL/10ScSn × DBA/2)F$_1$ mice kept at the Wright Fleming Institute at St Mary's Hospital Medical School, London. C57BL/10ScSn and CBA/Ca mice used as susceptible and resistant controls respectively for *Lsh* typing were obtained from Olac 1976 Ltd. Mice were typed individually for susceptibility to either *S. typhimurium* C5 or *L. donovani* infection as previously described[2,3]. In addition, (C57BL/6J × BXD-29)F$_1$ mice bred at the Jackson Laboratory were tested for resistance to *S. typhimurium* C5 (ref. 2) and TML (ref. 7) in the two laboratories.

Two BXD-18 and eight BXD-20 mice all typed as *Lsh*^{s} (Fig. 1). In each case, the original typing of *Lsh*^{r} by Bradley *et al.*[5] was based on examination of two mice whose parasite counts fell well within the *Lsh*^{r} range (Fig. 1, actual counts previously unpublished). This suggests that BXD-18 and BXD-20 had not proceeded to fixation for either allele at *Lsh* at that time. Our more recent *Lsh*^{s} typings are now consistent with the earlier[7] unequivocal *Ity*^{s} typings for BXD-18 and BXD-20 and thus eliminate two of the putative recombinants.

Lsh typing of eight BXD-28 and seven BXD-29 mice (Fig. 1) confirm the original typings (*Lsh*^{r} and *Lsh*^{s} respectively) of Bradley *et al.*[5]. In the case of BXD-28, the geometric mean spleen count of $4.61 + 0.98$ obtained after re-examining 20 mice with *S. typhimurium* C5 agrees closely with the mean of 4.46 obtained in the earlier[7] work with the TML strain of *S. typhimurium*. For BXD-29 mice, however, the geometric mean spleen count of $5.50 + 0.95$ based on examination of 34 mice with the C5 strain is significantly higher than the mean of 3.9 obtained after examination of only 5 mice with the TML strain, which led to the earlier[7] typing of BXD-29 as *Ity*^{r}. On the basis of our original criteria[4] for typing individual mice using 10^3 *S. typhimurium* C5 and day 10 organ counts, 22 of 34 BXD-29 mice with spleen and liver counts $>10^5$ type as *Ity*^{s}. The remainder were high in the *Ity*^{r} range with the geometric mean 2 logs higher than that for DBA/2 *Ity*^{r} controls. Our combined C5 and TML results (Fig. 2), which show 16 of 17 (C57BL/6J × BXD-29)F$_1$ and 26 of 28 (BXD-29 × C57BL)F$_1$ mice with spleen and liver counts clearly in the susceptible ($>10^5$) range, support the hypothesis that BXD-29 is, in fact, *Ity*^{s}. This would explain the previous finding[7] that 58% of BXD-29 mice tested died by day 30 when given only 25 organisms of the TML strain.

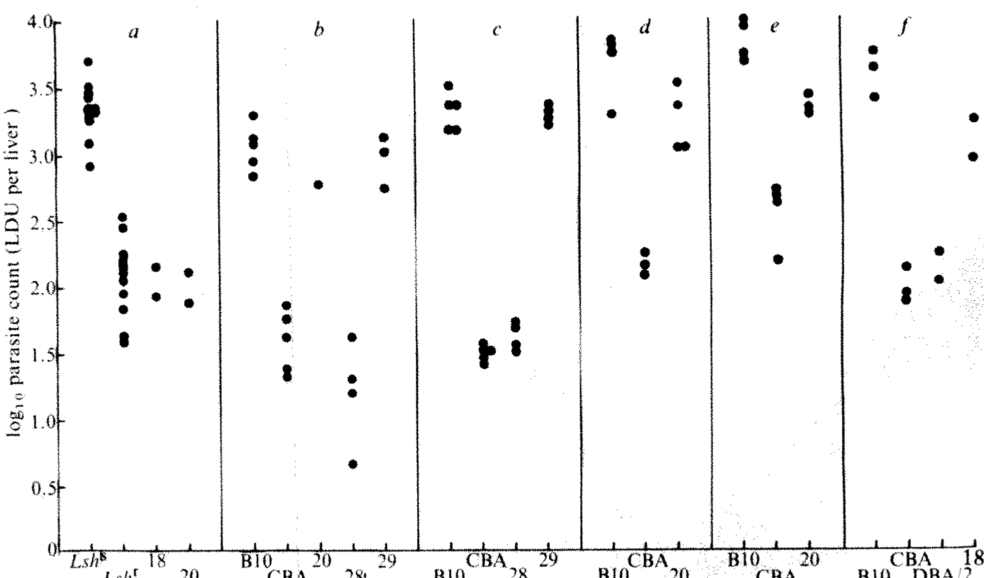


Fig. 1 *Lsh* typing experiments for RI (BXD-18, 20, 28, 29) and control *Lsh*^{s} (C57BL/10ScSn or B10) and *Lsh*^{r} (CBA/Ca and DBA/2) strains of mice. Mice were injected with $1-5 \times 10^7$ amastigotes of the L82 strain of *L. donovani* and examined 15 days later. The number of parasites per 500 liver cell nuclei was counted from Giemsa-stained impression smears and the results expressed as 'Leishman-Donovan units' (LDU) as described previously[12]. Counts for individual mice are shown for six separate typing experiments. *a* Shows actual counts for mice used in the original *Lsh* typing experiments of Bradley *et al.*[5]. Counts for BXD-18 and BXD-20 mice are compared with the range of *Lsh*^{s} and *Lsh*^{r} counts obtained for all BXD strains examined in that experiment. Both clearly fell well within the *Lsh*^{r} range of counts. *b–f* Show the results of our more recent typing experiments. Counts for BXD-28 and BXD-29 Mice (*b*, *c*) fell clearly within control *Lsh*^{r} and *Lsh*^{s} ranges respectively. Counts for BXD-18 and BXD-20 mice (*b*, *d–f*) always fell towards the lower end of the *Lsh*^{s} range of counts, irrespective of parasite dose. Although this difference was not always significant in individual experiments, it does suggest some additional modifying genetic control of *Lsh*^{s} counts. B10, BXD-18 and BXD-20 counts all differed significantly from control *Lsh*^{r} counts ($P < 0.001$) and, in each experiment, there was at least a 0.5 log$_{10}$ LDU separation between the lowest susceptible and highest resistant counts. The retyping of BXD-18 and BXD-20 as *Lsh*^{s} has recently been confirmed in another laboratory[10] where parasite counts again fell at the lower end of the susceptible range of counts.

0028-0836/82/230510-02$01.00

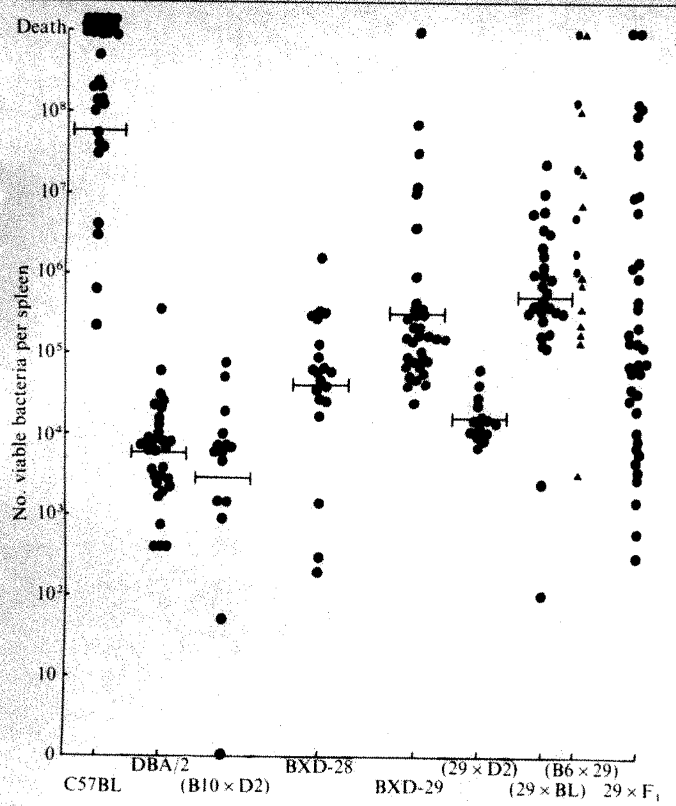


Fig. 2 *Ity* typing experiments for RI (BXD-28, 29) strains and the F_1 hybrid combinations: (C57BL/10ScSn × DBA/2) or (B10 × D2); (BXD-29 × DBA/2) or (29 × D2); (BXD-29 × C57BL) or (29 × BL); (C57BL/6J × BXD-29) or (B6 × 29); and [BXD-29 × (C57BL/10ScSn × DBA/2)F_1] or (29 × F_1). Mice were injected subcutaneously with 10^3 *S. typhimurium* strain C5 (●) or TML (▲) and examined 10 or 9 days later respectively. Viable bacteria in spleens and livers were estimated as previously described[2,4,7]. Bacterial loads in the livers agreed closely with the viable bacterial counts per spleen shown for individual mice. Dead mice were estimated at 10^8 bacteria per spleen in calculating the $\log_{10}$ geometric mean spleen counts indicated for each group. The RI strains and hybrids are compared with control Ity^s (C57BL) and Ity^r (DBA/2) strains.

A range of sensitivity from highly susceptible (10^7–10^8 counts or death by day 10) through intermediate (10^5–10^6) to highly resistant (10^2–10^4) was observed in [BXD-29 × (C57BL/10ScSn × DBA/2)F_1] backcross mice (Fig. 2). This implies multigenic control additional to *Ity*. The possibility of additional genetic influence on the sensitivity to *S. typhimurium* infection in the intermediate BXD strains has already been considered[7] and is consistent with the previous observations[1,8,9] of increased susceptibility in the Ity^r DBA/2 progenitor strain compared with other Ity^r strains. A single additional gene bearing susceptible alleles in the DBA/2 progenitor and resistant alleles in the C57BL/6 progenitor is insufficient, however, to explain the present results. We propose, therefore, that at least two other genes are involved in modifying the phenotypic expression of *Ity* in the BXD intermediate strains.

We therefore conclude that it is not proved that *Ity* and *Lsh* are distinct genetic loci. The fact that no recombinants were detected among 22 RI strains from 4 other progenitor strain combinations[7] supports our hypothesis that the progenitors of BXD are introducing additional genetic influence on the outcome of salmonella infections. The recent finding[10] that *Bcg*, a gene controlling innate resistance to *Mycobacterium bovis* infection in mice,[11] types in concordance with the present *Lsh* and *Ity* typings for the BXD strains also supports our conclusions. We therefore favour the hypothesis that a single gene on chromosome 1 controls innate resistance and susceptibility to these three, and possibly other, intracellular infections via one mechanism, rather than a complex of tightly linked genes.

We thank Jane Ferguson and Orysia Ulczak for assistance in the laboratory, Ben Taylor for supplying the RI strains and the Wellcome Trust for financing trans-Atlantic discussion. The work was supported by grants from the MRC, the Wellcome Foundation and the Uniformed Services University of the Health Sciences.

Received 30 July 1981; accepted 10 March 1982.

1. Plant, J. & Glynn, A. A. *Nature* **248**, 345–347 (1974).
2. Plant, J. & Glynn, A. A. *J. infect. Dis.* **133**, 72–78 (1976).
3. Bradley, D. J. *Clin. exp. Immun.* **30**, 130–140 (1977).
4. Plant, J. & Glynn, A. A. *Clin. exp. Immun.* **37**, 1–6 (1979).
5. Bradley, D. J., Taylor, B. A., Blackwell, J., Evans, E. P. & Freeman, J. *Clin. exp. Immun.* **37**, 7–14 (1979).
6. Bradley, D. J. *Nature* **250**, 353 (1974).
7. O'Brien, A. D., Rosenstreich, D. L. & Taylor, B. A. *Nature* **287**, 440–442 (1980).
8. Hormaeche, C. E. *Immunology* **37**, 319–327 (1979).
9. Benjamin, W. H. & Briles, D. E. *Abstr. an. Meet. Am. Soc. for Microbiol.* No. E51, 63, Texas (1981).
10. Skamene, E. *et al. Nature* **297**, 506–509 (1982).
11. Gros, P., Skamene, E. & Forget, A. *J. Immun.* **127**, 2417–2421 (1981).
12. Bradley, D. J. & Kirkley, *J. Clin. exp. Immun.* **30**, 119–129 (1979).

Recognition of conformational determinants on H–2 by cytolytic T lymphocytes

Linda A. Sherman

Department of Immunopathology, Scripps Clinic & Research Foundation, La Jolla, California 92037, USA

The specificity of cytolytic T lymphocytes (CTL) differs from that of B lymphocytes insofar as they exhibit a marked preference for recognition of major histocompatibility complex (MHC) encoded cell surface antigens. In murine systems, this is reflected both in the high frequency of CTL which recognize H–2 alloantigens[1–5] and also in the requirement for recognition of syngeneic H–2 antigens in conjunction with cellularly presented foreign antigens, so called MHC restricted recognition[6–10]. Various models for T-cell recognition have been proposed in an attempt to explain both this predisposition for MHC recognition and the basis for restricted recognition (reviewed in refs 10, 11). Most models have fallen into one of two general categories; those which attribute recognition to two independent combining sites on the T cell receptor(s), one specific for foreign antigens and one specific for MHC antigens; and those which accommodate recognition of both molecules using a single combining site which recognizes a neoantigenic determinant composed of a complex of both molecules. In order to understand the basis for determinant recognition, the specificity of individual clones of C57BL/6 CTL induced in response to a point mutation in the H–2K^b molecule has been analysed. Using a series of H–2K^b mutants as targets it was found that a single clone can recognize two or more mutants which, although different from each other with respect to the position of their amino acid substitutions, appear to have independently gained the same new determinant. This strongly suggests that CTL can respond to conformational nuances in self H–2 molecules as shared by these different K^b mutants. On this basis, it is proposed that MHC restricted recognition may represent the recognition of conformational alterations of self resulting from the interaction of MHC molecules and foreign antigens.

In order to approach questions concerning T-cell recognition, this laboratory has carried out a detailed analysis of the recognition of H–2 antigens by individual CTL clones[12–15]. Most revealing in terms of probing the basis for determinant recognition have been the results of fine specificity analysis of monoclonal CTL stimulated as a result of a single amino acid substitution in the H–2K^b molecule, the B6.C–H–2^{bm11} anti-C57BL/6 response (bm11 anti-B6)[13]. A large proportion (25%) of such clones did not recognize any of the five different H–2K^b mutants which were used as targets. Considering that each of these mutants differed from the stimulating K^b molecule by only one

or two amino acid substitutions, and that in most cases each mutation occurred at different positions on the molecule[16], it is unlikely that all of these different mutations could have resulted in an amino acid substitution in the particular determinant recognized by a single clone. A more likely explanation for such lack of recognition would be that, in some if not all mutants, these particular wild-type determinants have been lost due to conformational changes resulting from the mutation[17,18]. In the case of antibody recognition of globular proteins, such a mechanism has previously been used to explain how the antigenic integrity of a particular determinant can be disrupted by a distal amino acid substitution[19].

This suggests the possibility that a conformational alteration might also create new determinants which were, by analogy, in positions on the molecule different from the site of the amino acid substitution responsible for their creation. In support of this possibility, Melief and co-workers previously demonstrated that, although B6-anti-mutant CTL populations would most strongly react against the K^b mutant used as stimulator these CTL demonstrated detectable cross-reactivity on a variety of disparate K^b mutants[18]. This suggested different K^b mutants may have gained some similar determinants. A possible mechanism would be that each mutation creates its own set of conformational differences in the K^b molecule and that amino acid substitutions in different parts of the molecule could independently effect a conformational change which, fortuitously, resulted in a similar determinant.

The experiments described in this report constitute a test of this hypothesis. CTL clones of wild-type origin were stimulated in primary limiting dilution cultures against the bm11 mutant. It was reasoned that the wild-type responder would be tolerant of all K^b determinants retained by the bm11 mutant and, therefore, such clones should specifically recognize new determinants attributable to the bm11 mutation. Each clone was tested for cross-reactive recognition on a panel of targets which included the wild type B6, six other K^b mutants, and the allogeneic D2.GD target (K^d, D^b). Where known, the positions of the amino acid substitutions in each of these mutants is as follows: bm1, 155, and 156; bm3, 77 and 89; bm8, 23; bm9, 116 and 121; bm10, 165; and bm11, 77 (ref. 16).

Results obtained from panel analysis of 58 B6-anti-bm11 clones independently derived from 11 individual B6 responders are presented in Table 1. Positive recognition refers to cross-reactive lysis which represents at least 60% of the value obtained on the bm11 target as indicated in the legend to Table 1. As expected, none of the B6 clones demonstrated specific recognition of the syngeneic B6 targets. Also, as expected on the basis of a shared mutation at amino acid position 77, the vast majority of bm11 specific clones (83%) also recognized the bm3 target. It is likely that the antigenic determinants recognized by the 17% which did not recognize bm3 are obscured or disrupted on the bm3 molecule as a result of the second mutation at position 89. Most interesting, however, is the observation that the majority of clones (62%) recognized not only targets which bear the mutation in position 77, but also recognize at least one of the other K^b mutants such as bm8, bm1 and bm9 which are identical to the wild type K^b in position 77, having their substitutions elsewhere. Moreover, of these 26 clones, 24% recognize at least two such H-2K^b mutants.

In Table 2, these data are presented as the fraction of clones that recognize each of these targets. The fraction of clones that cross-react on the third-party alloantigen, K^d on D2.GD, is comparable to the level of third-party reactivity usually seen in conventional allogeneic responses[3-5] (5–10%). In contrast, some of the mutants, bm8, bm1, bm9, are recognized by a significantly greater proportion of clones than is usual for third party recognition. Based on these results, it may be concluded that different and often distal amino acid substitutions in an H-2 molecule can fortuitously result in antigenic determinants which if not identical, are at least similar enough to permit comparable T-cell recognition as reflected in the comparable degree of lysis of different mutant targets.

In view of the disparate positions of most of these mutations, it would appear unlikely that two different mutants could demonstrate significant cross-reactivity on the basis of a common 'foreign' determinant. The only candidates for a non-self antigen as perceived by the strain of origin would be determinants that encompass the substituted residues and it is unlikely these are cross-reactive since they would correspond to a different position on the molecule for each mutant. Therefore, these results strongly support the hypothesis proposed above, that conformational alterations of self H-2 antigens can result in immunogenic determinants which do not include the

Table 1 Panel analysis of B6 anti-bm11 CTL clones

Reactivity patterns	C57BL/6Kh	bm1	bm3	bm4	bm8	bm9	bm10	bm11	D2.Gd	No. of clones	% Of total
a	·	·	·	·	·	·	·	+	·	6	10.3
b	·	+	·	·	·	·	·	+	·	1	1.7
c	·	·	+	·	·	·	·	+	·	13	22.4
d	·	·	·	·	+	·	·	+	·	2	3.4
e	·	·	·	·	·	+	·	+	·	1	1.7
f	·	+	+	·	·	·	·	+	·	9	15.5
g	·	+	+	+	·	·	·	+	·	2	3.4
h	·	·	+	+	·	·	·	+	·	3	5.2
i	·	·	+	·	+	·	·	+	·	6	10.3
j	·	·	+	·	+	+	·	+	·	1	1.7
k	·	·	+	·	·	+	+	+	·	2	3.4
l	·	+	+	·	·	+	+	+	·	1	1.7
m	·	·	+	·	·	+	·	+	·	7	12.1
n	·	·	+	·	·	·	·	+	+	2	3.4
o	·	+	+	·	·	·	·	+	+	1	1.7
p	·	·	+	·	+	·	·	+	+	1	1.7

B6 splenic CTL precursors were stimulated *in vitro* under limiting dilution culture conditions in V bottom microtitre wells containing 5×10^5 irradiated B6 'helper' spleen cells, 5×10^6 irradiated bm11 stimulator cells, and varying numbers (1,000–6,000) of B6 responder cells in 0.2 ml of culture media as previously described[12]. After 6 days in culture, a 60 μl portion of each well was tested for cytolytic activity against conA stimulated bm11 targets. Such target cells were prepared and labelled with ^{125}I-iododeoxyuridine[12] and 1×10^4 cells were added to each well in 140 μl of culture media containing 10% fetal bovine serum and 3% rat-T-cell growth factor which was a necessary addition to maintain maximum viability of these target cells during long assays (18–20 h). CTL clones obtained using responder cell numbers resulting in 20% or fewer positive wells were expanded[12] and tested for lytic activity against ^{51}Cr-labelled bm11 target cells. Successfully expanded clones were assayed in duplicate on each of the indicated target cells. A clone was considered positive or negative for recognition of a particular target if its value for lysis was either >60% or <25% respectively of its value for lysis on bm11 targets. Clones which did not meet these criteria on all targets were considered ambiguous and not recorded above.

Table 2 Cross-reactive recognition of H–2K^b mutants by B6 anti-bm11 CTL clones

Strain	% Clones
bm11	100
bm3	83
bm1	24
bm9	21
bm8	17
bm4	9
bm10	5
D2.GD	7

amino acid substitution responsible for the alteration.

Based on these results, it is further proposed that the interaction of foreign antigens with H–2 encoded molecules can similarly result in conformational changes which may be immunogenic. A similar hypothesis has recently been proposed by Ohno[20]. In order to conform with our views of restricted recognition, such interaction would need to fulfil the following requirements: (1) the same antigen should always result in the same conformation change in H–2, and (2) this interaction should be sufficiently stable to permit binding and triggering of T receptors that recognize the resultant novel H–2 conformation. It would then be possible to explain the appearance of antigen specificity and MHC restriction even though the receptor actually recognized 'altered self' and not antigen.

There are several observations which are consistent with such a model. First, there is much evidence for stability of H–2 antigen interactions as measured either by co-capping or co-purification of these two molecules[21–25]. Second, it has proven much easier to block antigen specific CTL or MLC using anti-H-2 antibody reagents as opposed to using antibody specific for the foreign antigen *per se*[26–29]. This is most consistent with an indirect role for antigen in receptor recognition. Third, it has been reported by several laboratories that influenza specific CTL are more cross-reactive than influenza specific antibody[30–32]. This is again most readily accommodated by a model which predicts that only some foreign determinants potentially contribute to the specificity of a CTL response, in particular, those which effect the interaction between H–2 and the foreign molecule. Finally, this model presents a structural basis for the well documented cross-reactivity against allo-antigen of antigen specific T cells, since both types of recognition would involve altered-self MHC determinants[33–36].

The implication of this hypothesis which may be most difficult to accept is the burden placed upon the H–2 molecule by the fact that the diversity of all possible H–2 determinants would have to be sufficient to account for antigen specificity. However, if the low frequency of CTL precursors specific for non-MHC antigens, such as virus and minor antigens[5,37], reflects a correspondingly small number of different determinants created by an antigen H–2 interaction, then the chances that a second antigen would create the same determinants may be relatively slim. In contrast, in essentially all responses where the H–2 antigen is itself structurally altered, many new determinants are created, thereby greatly increasing the likelihood of cross reactivity. This would account for the observation presented in this report.

These studies were supported by NIH grant nos AI 15710 and CA 25803.

Received 22 December 1981; accepted 30 March 1982.

1. Wilson, D. B., Blythe, J. & Nowell, P. C. *J. exp. Med.* **128**, 1157–1181 (1968).
2. Bevan, M. J., Langman, R. E. & Cohn, M. *Eur. J. Immun.* **6**, 150–156 (1976).
3. Lindhal, K. F. & Wilson, D. B. *J. exp. Med.* **145**, 508–522 (1977).
4. Miller, R. G., Teh, H.-S., Harley, E. & Philipps, R. A. *Immun. Rev.* **35**, 38–58 (1977).
5. MacDonald, H. R. *et al. Immun. Rev.* **51**, 93–123 (1980).
6. Zinkernagel, R. M. & Doherty, P. C. *Nature* **248**, 701–702 (1974).
7. Shearer, G. M. *Eur. J. Immun.* **4**, 527–533 (1974).
8. Bevan, M. J. *J. exp. Med.* **142**, 1349–1364 (1975).
9. Forman, J. *J. exp. Med.* **142**, 403–418 (1975).
10. Zinkernagel, R. M. & Doherty, P. C. *Adv. Immun.* **27**, 51–177 (1979).
11. Matzinger, P. *Nature* **292**, 497–501 (1981).
12. Sherman, L. A. *J. exp. Med.* **151**, 1386–1397 (1980).
13. Sherman, L. A. *J. Immun.* **127**, 1259–1260 (1981).
14. Sherman, L. A. *J. exp. Med.* **155**, 380 (1982).
15. Sherman, L. A. *Isolation, Characterization and Utilization of T. Lymphocytes* (eds Fathman, C. G. & Fitch, F.) (Academic, New York, in the press).
16. Nairn, R., Yamaga, K. & Nathenson, S. G. *A. Rev. Genet.* **14**, 241–277 (1980).
17. Klein, J. *Adv. Immun.* **26**, 55–146 (1978).
18. Melief, C. J. M., de Waal, L. P., van der Meulen, M. Y., Melvold, R. W. & Kohn, H. L. *J. exp. Med.* **151**, 993–1013 (1980).
19. White, T. J., Ibrahimi, I. M. & Wilson, A. C. *Nature* **274**, 92–94 (1978).
20. Ohno, S., Epplen, J. T. & Sutou, S. *Hum. Genet.* **58**, 37–45 (1981).
21. Schrader, J. W., Cunningham, B. A. & Edelman, G. M. *Proc. natn. Acad. Sci. U.S.A.* **72**, 5066–5070 (1975).
22. Senik, A., Demant, P. & Neauport-Sautes, C. *J. Immun.* **122**, 1461–1467 (1979).
23. Zarling, D. A., Keshet, I., Watson, A. & Bach, F. *Scand. J. Immun.* **8**, 497–508 (1978).
24. Bubbers, E. J., Chen, S. & Lilly, F. *J. exp. Med.* **147**, 340–351 (1978).
25. Honneycutt, P. J. & Gooding, L. R. *Eur. J. Immun.* **10**, 363–370 (1980).
26. Blanden R. V., Hapel, A. J., Doherty, P. C. & Zinkernagel, R. M. in *Immunobiology of the Macrophage* (ed. Nelson, D. S.) 367–400 (Academic, New York, 1976).
27. Braciale, T. J. *Cell. Immun.* **33**, 423–436 (1977).
28. Doherty, P. C., Blanden R. V. & Zinkernagel, R. M. *Transplant. Rev.* **29**, 89–124 (1976).
29. Askonas, B. A. & Webster, R. G. *Eur. J. Immun.* **10**, 151–156 (1980).
30. Effros, R. B., Doherty, P. C., Gerhard, W. & Bennink, J. *J. exp. Med.* **145**, 557–568.
31. Braciale, T. J. *J. exp. Med.* **149**, 856–869 (1979).
32. Ada, G. L. & Yap, K. L. *Immunochemistry* **14**, 643–651 (1977).
33. Finberg, R., Burakoff, S., Cantor, H., & Benacerraf, B. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5145–5149 (1978).
34. von Boehmer, H. *et al. Eur. J. Immun.* **9**, 592–597 (1979).
35. Sredni, B. & Schwartz, R. H. *Nature* **287**, 855–857 (1980).
36. Braciale, T. J., Andrew, M. E. & Braciale, V. L. *J. exp. Med.* **153**, 1371–1376 (1981).
37. Wagner, H. *et al. Immun. Rev.* **58**, 95–129 (1981).

Cytogenetic mapping of the duplicated segment of chromosome 12 in lymphoproliferative disorders

G. Gahrton*†, K.-H. Robert*, K. Friberg†, G. Juliusson*, P. Biberfeld† & L. Zech†

* Section of Clinical Hematology and Oncology, Department of Medicine, Huddinge Hospital, S-14186 Huddinge, Sweden
† Institute of Medical Cell Genetics, Medical Nobel Institute and Department of Pathology, Karolinska Institute, S-10401, Stockholm, Sweden

Recently, an extra chromosome 12 has been reported to occur with high frequency in peripheral blood lymphocytes from patients with chronic B-cell lymphocytic leukaemia (B-CLL)[1,2]. Here we report a unique chromosomal abnormality, which suggests that the segment q13 to q22 on chromosome 12 carries the important genes that are duplicated in those lymphoproliferative disorders characterized by trisomy for this chromosome.

Cytogenetic studies were performed on peripheral blood lymphocytes, bone marrow cells and spleen cells from a 56 year-old man with a lymphocytic non-Hodgkin lymphoma. The lymphoma was found in the spleen, bone marrow, and peripheral blood. The white blood cell count was 9.3×10^9 per litre and a differential count showed that 74% of the cells were lymphocytes. Surface membrane immunofluorescence studies revealed a monoclonal B-cell proliferation with a phenotype $\mu\kappa$ when tested with fluorescein-conjugated $F(ab')_2$ fragments of goat anti-human μ, γ, δ, κ, and λ antisera. Microscopy of the spleen and small lymph nodes showed a mixed lymphocytic–histiocytic lymphoma (Rappaport classification[3]) or polymorphic immunocytoma (Kiel classification[4]).

Table 1 Metaphase analysis of lymphocytes from a patient with chronic B-cell lymphocytic leukaemia

Material	Mitogen	No. of metaphases analysed	No. of abnormal metaphases	Karyotype
Peripheral blood	EBV	17	2	46XY;
	LPS	11	3	46XY, t(10;10)
	PHA	21	1	(pter; q24),
Bone marrow	EBV	6	2	dup(12)
Spleen	EBV	5	4	(q13—q22)
	LPS	3	1	
Total		63	13	

0028-0836/82/230513—02$01.00

Mitogenic stimulation of isolated peripheral blood lymphocytes was performed with Epstein–Barr virus (EBV), lipopolysaccharide (LPS) from *Escherichia coli*, and phytohaemagglutinin (PHA)[1]. Cultures for cytogenetic analysis were collected on day 4. Metaphase preparations were prepared by conventional methods and the preparations stained according to the Q-banding technique[5]. The results of the cytogenetic analyses are shown in Fig. 1 and Table 1. Of 63 metaphases, 13 had the same chromosomal abnormality. One chromosome 12 had an extra part attached at band 22, which was identical with the part q13 → qter of a normal chromosome 12. The other chromosome 12 was normal. Thus, the region q13 → q22 was duplicated. In addition to this abnormality, a balanced reciprocal translocation between the two chromosomes 10 was found in all metaphases carrying the abnormal chromosome 12.

As shown in Fig. 2, the observed abnormality of chromosome 12 has probably arisen through chromatid exchange. It is probable that a breakpoint occurred on one chromatid of one chromosome 12 at q13 and another breakpoint on the other chromosome 12 at q22. The two chromatids were then exchanged. Alternatively, there may have been three breakpoints, two on one chromatid of one chromosome 12 at q13 and q22, and one on q13 or q22 on one chromatid of the other chromosome 12. The chromatid part, q13—q22, on the first chromosome was then inserted at the breakpoint of the second chromosone 12. Of the possible abnormal karyotypes that can appear, irrespective of the mechanism, only alternative *b* in Fig. 2 was found. Normal karyotypes were also found, but these were probably due to the presence of normal nonmalignant lymphocytes in the peripheral blood[1].

The results show that the essential genes that tend to be duplicated during leukaemogenesis leading to lymphoproliferative B-cell disorders characterized by trisomy of chromosome 12, are probably located on the seqment q13 → q22.

Other experimental work on murine lymphomas has also indicated that only part of a trisomic chromosome is crucial for the neoplastic transformation[6]. Mouse T-cell lymphomas are characterized by triosmy 15, but only genes located distal to band T6 are important for leukaemogenesis. Furthermore, in certain human myeloproliferative disorders characterized by trisomy of chromosome 1, there are indications that only the segments q23—q25[7,8] or q25—q32[9] are associated with disease. Thus, in lymphoproliferative disorders (and probably also other haematopoietic diseases) characterized by trisomic chromosomes, the important genes to be duplicated for disease development are probably located on a very small, specific part of the chromosome. In human lymphoproliferative disorders characterized by trisomy 12, these genes seem to be located on the segment q13—q22. So far mapping of chromosome 12 has only located the loci for a few enzymes and antigens to the long arm of this chromosome, that is mitochondrial synthetase, serine hydroxymethyltransferase, branched-chain amino acid transferase-1, and surface antigen 8 (ref. 10). Further studies of these enzymes in lymphoproliferative B-cell disorders may elucidate whether any part of the q13—q22 region on chromosome 12 is also involved in patients where no gross cytogenetic abnormalities are found.

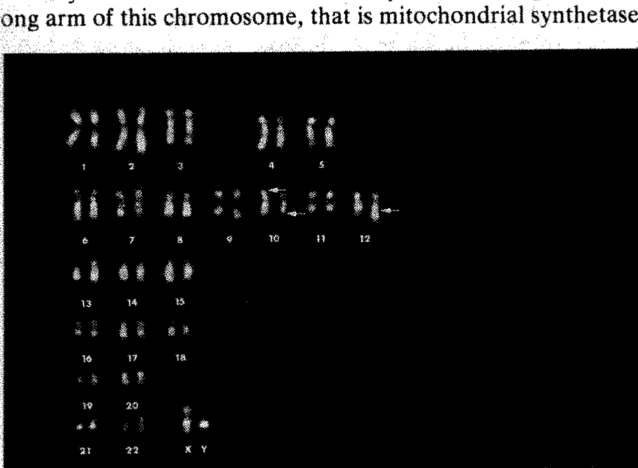


Fig. 1 Karyotype of a peripheral blood lymphocyte stimulated by EBV, showing duplication of the seqment q13–q22 on chromosome 12. There is also a reciprocal balanced translocation between the two chromosomes 10.

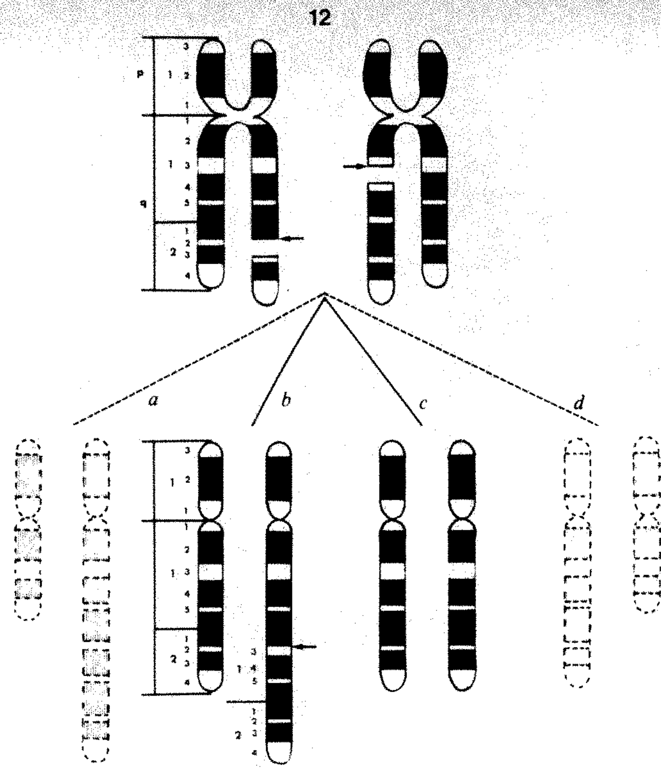


Fig. 2 Possible mechanism of development of the chromosomal abnormality. Only alternatives *b* and *c* were found.

Received 8 February; accepted 15 April 1982.

1. Gahrton, G., Robèrt, K.-H., Friberg, K., Zech, L. & Bird, A. G. *Blood* **56**, 640–647, (1980).
2. Gahrton, G., Robèrt, K.-H. & Zech, L. *Blood* **58**, 859 (1981).
3. Rappaport, H., Winter, W. J. & Hicks, E. B. *Cancer* **9**, 792–821 (1956).
4. Lennert, K. & Mohri, N. *Malignant Lymphomas other than Hodgkin's Disease*, 111–469 (Springer, New York, 1978).
5. Caspersson, T., Lomakka, T. & Zech, L. *Hereditas* **67**, 89–102 (1971).
6. Wiener, F., Ohno, S., Spira, J., Heran–Ghera, N & Klein, G. *Nature* **257**, 658–660 (1978).
7. Gahrton, G., Friberg, K., Lindsten, L. & Zech, L. *Hereditas* **88**, 1–5 (1978).
8. Gahrton, G., Friberg, K., Zech, L. & Lindsten, J. *Lancet* **i**, 96–97 (1978).
9. Rowley, J. D. *Proc. natn. Acad. Sci U.S.A.* **74**, 5792 (1977).
10. McKusik, V. A. *Cytogenet. cell Genet.* (in the press).

Two distinct forms of surface antigen gene rearrangement in *Trypanosoma brucei*

Phelix A. O. Majiwa, John R. Young, Paul T. Englund*, Stuart Z. Shapiro & Richard O. Williams

International Laboratory for Research on Animal Diseases, PO Box 30709, Nairobi, Kenya.
*Department of Physiological Chemistry, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

African trypanosomes express different surface antigens sequentially in their mammalian hosts[1]. Expression of different variable surface glycoprotein (VSG) antigens is associated with rearrangements of genomic DNA[2-5]. Several laboratories have reported that expression of some VSG genes is accompanied by the appearance of a duplicated copy of the genes at a new location in the genome[3,4]. However, in a series of sequentially related *Trypanosoma brucei* clones from a different stock we were unable to detect duplication of one VSG gene (ILTAT 1.2, B) in expressing clones. Two copies of this gene were found

in expressing and non-expressing clones. Apparent insertions and deletions occurred near the 3′ ends of both gene copies, but could not be correlated with expression[2,5,6]. We report here the observation of both types of rearrangement for different VSG genes in the same series of trypanosome clones. Thus the different types of rearrangement are not a function of different trypanosome stocks, but of different VSG genes within a stock.

The trypanosome clones used in this study were derived from a single clone (clone A, expressing VSG A) from *T. brucei* stock 227 (ref. 7). Clone B was derived from a relapse infection of clone A and clones C and D were cloned from the first-relapse populations in normal mice infected with clone B. Trypanosomes grown in lethally irradiated rats were checked by immunofluorescence for homogeneity of VSG expression[7] before isolation of RNA or DNA.

The construction and identification of plasmids pcBB1 and pcBC1 containing mRNA sequences encoding VSGs B (ILTAT 1.2) and C (ILTAT 1.3), respectively, have been described elsewhere[8]. Plasmid pcBA1 was constructed and initially identified in a similar manner. That pcBA1 contains VSG A coding sequences is substantiated by two further criteria: it contains 3′-coding and non-coding sequence homologies with other VSG mRNAs[9], and hybridizes to RNA found only in trypanosome clone A. This is shown in Fig. 1, where pcBA1 was hybridized to total RNA from trypanosome clones expressing VSGs A, B, C and D (ILTAT 1.4). The inserted sequences of pcBA1 are present only in the trypanosome clone expressing VSG A. The only known clone-specific mRNA is that encoding the VSG expressed[10]. In addition to the major band at 2 kilobases (kb), the expected size of mature VSG mRNA, pcBA1 also hybridizes to three faint bands of 3.2, 3.6 and 4 kb, in clone A RNA only. As electrophoresis was done in stronger denaturing conditions[11], these bands may represent incompletely processed precursors of the VSG A mRNA. Such precursor RNAs have not been described previously for other VSGs.

The experiment shown in Fig. 2 characterizes the rearrangements affecting the genes for VSGs A, B, and C. Complementary DNA sequences encoding each of the VSGs

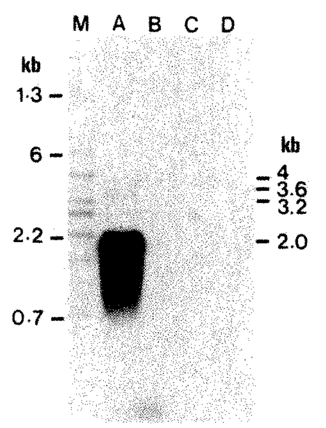


Fig. 1 Clone specificity of the cDNA sequence in pcBA1. Total RNA from four trypanosome populations homogeneously expressing different VSGs (A, B, C and D) was prepared by dissolving 10^{10} trypanosomes in 10 ml of 60% guanidinium thiocyanate, 50 mM sodium citrate *p*H 7.0, 0.5% sarcosyl. This was layered over 1.5 ml cushions of 5.7 M caesium chloride containing 50 mM EDTA *p*H 7.5, and centrifuged at 30,000 r.p.m. for 6 h in a Bechman SW 50.1 rotor. DNA remained at the interface and RNA was recovered from the pellet after careful removal of the supernatant and cushion; ~2.5 μg from each clone was denatured with glyoxal, electrophoresed, transferred to nitrocellulose, and hybridized with radioactive pcBA1 as described by Thomas[11]. The filter was washed twice in 0.1×SSC, 0.1% SDS at 50 °C for 15 min before autoradiography. The mRNA band was deliberately over-exposed to show the complete specificity of hybridization at this stringency, and to reveal the three faint high molecular weight bands in clone A. Molecular weight markers are shown under column M.

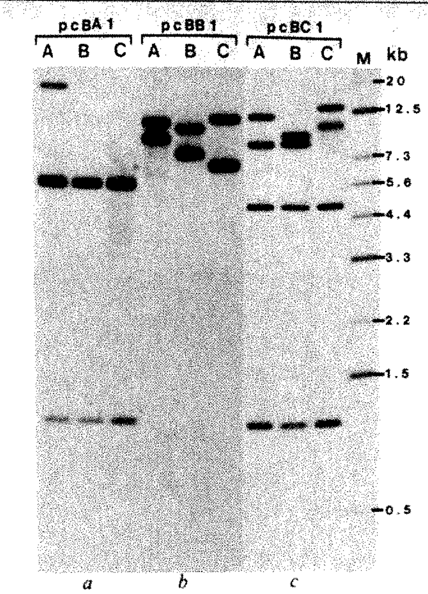


Fig. 2 Comparison of rearrangements affecting the genes for VSGs A (ILTAT 1.1), B (ILTAT 1.2) and C (ILTAT 1.3). DNA from trypanosome clones expressing each of these VSGs was digested with *Hin*cII (for hybridization to pcBA1 and to pcBC1) and *Sal*I (for hybridization to pcBB1). These enzymes do not have cleavage sites in the cDNAs to which their digests were hybridized. Fragments were separated by electrophoresis on a single gel and transferred to nitrocellulose in 20×SSC as described elsewhere[14]. Cut sections of the filter were hybridized, in conditions described elsewhere[14], to plasmids pcBA1 (*a*), pcBB1 (*b*) and 1,100-bp 5′ fragment of pcBC1 (*c*), which had been labelled with ^{32}P by the procedure of Rigby *et al.*[15]. The markers (M) are fragments of known size which hybridize to pBR322 sequences in the pcBC1 probe. The filter sections were washed twice in 0.1×SSC, 0.1% SDS at 65 °C for 1 h, dried, and then realigned for exposure to X-ray film.

were hybridized to DNA from trypanosome clones expressing the three antigens. The enzyme used in each case did not cut the cDNA sequence of the VSG probe used. Thus, if there are no introns in the gene, each hybridizing band visualized implies a distinct region that is homologous with the probe.

For pcBA1, hybridizing fragments were identical in size in each clone except for the presence of an extra, larger fragment in DNA from clone A. A different result was seen for the other two probes. There was no additional fragment in the DNA from the cell clones expressing VSG B or C, as there was neither an extra band nor an increase in relative intensity of any band. Although the number of fragments hybridizing to each cDNA was the same for all trypanosome clones, there were large differences in the sizes of two of the hybridizing fragments, even when DNA from two non-expressing clones was compared.

Gene rearrangements were observed for all three cDNAs studied. Where an extra gene fragment was detected in the clone expressing the VSG encoded by the probe sequence (A), the remaining copies of the gene did not undergo rearrangements in different clones (Fig. 2). These characteristics of the VSG A gene rearrangement parallel observations by others using different VSG genes which are duplicated and transposed when expressed[3,4].

The experiment shown in Fig. 3*a*, using another enzyme that does not cut the cDNA, also demonstrates an extra fragment hybridizing to pcBA1 in clone A. This fragment hybridizes strongly to probe fragments from either end of the cDNA, indicating that it contains both ends of the cDNA sequence.

For other VSG genes which are duplicated and transposed when expressed, the duplicated segment extends from a point within the 3′ end of the mRNA sequence to a point beyond its 5′ end[12]. When enzymes cutting inside a transposed segment were used to observe duplicated genes, single bands were visualized with increased intensity. For VSG A, this was demonstrated

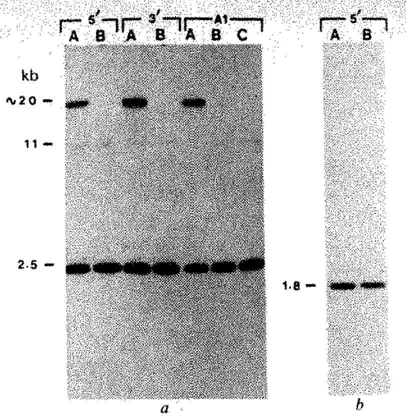


Fig. 3 Confirmation that the extra fragment hybridizing to pcBA1 contains intact most of the VSG A mRNA sequence. *a*, Hybridization of 5' and 3' fragments of pcBA1 to *Eco*RI digests of DNA from trypanosome clones expressing (A) and not expressing (B, C) VSG A, compared with hybridization of the whole cDNA. The 5' fragment contains 150 base pairs (bp) of the 900-bp cDNA sequence; the 3' fragment contains 380-bp, and includes part of the poly(A) tail. Both ends of the cDNA sequence are detected in the extra *Eco*RI fragment in clone A. *b*, Hybridization of the same 5' probe to *Pvu*II-digested DNA from clones A and B. The same *Pvu*II fragment, which contains most of the VSG A mRNA sequence, is produced from the extra copy unique to clone A and the copy present in non-expressing clone. Restriction enzyme fragments of pcBA1 were isolated by the procedure of Maxam and Gilbert[16]. Labelling of probes, electrophoresis and filter hybridization are described in Fig. 2 legend.

with enzyme *Pvu*II, (Fig. 3*b*). In clone A, an increase in intensity was observed in a 1.8-kb enzyme fragment, which was produced by the *Pvu*II site in the cDNA and another 1.8 kb towards the 5' end of the mRNA. Thus we suggest that the duplicated copy of the VSG A coding sequence is an intact copy of the cDNA sequence. It is reasonable to conclude that we are observing the same kind of duplication–transposition event as reported for other VSG genes.

Figure 4 shows that there are sequence differences between the basic copy and the duplicated extra copy of the clone A gene. We know from the published sequence of the clone A cDNA[9] that the coding sequence for the A antigen is not cut by *Bam*HI. As seen in Fig. 4, the extra copy of the clone A gene was not cut, whereas the basic copy in all cell clones was cut by this enzyme. A comparison of the cDNA sequence with the basic copy sequence showed that a point mutation has occurred at base position 1,700, resulting in the loss of a *Bam*HI site in the cDNA and therefore presumably in the duplicated extra copy of clone A (see ref. 13).

In previous descriptions of this type of gene rearrangement for the VSG B genes[2,5,6], we suggested that the differences between our observations with this gene, and those of others

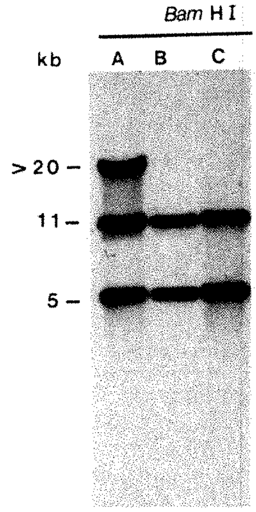


Fig. 4 Restriction enzyme site difference between the extra fragment and the basic copy of the clone A gene. pcBA1 was hybridized with *Bam*HI digests of DNA from expressing (A) and non-expressing (B, C) trypanosome clones. The probe recognizes two fragments (11 and 5 kb) in all clones and a fragment >20 kb in the expressing clone alone. When used as probes, the fragments from extreme ends of the cDNA (as in Fig. 3) show that the 11 and 5 kb fragments are respectively 5' and 3' portions of the basic copy (data not shown). This indicates the presence in the basic copy of a *Bam*HI site which is absent from both the cDNA and the additional, larger fragment unique to the expressor.

with other VSG genes, might be due to the use of different trypanosome stocks. This explanation cannot hold, as both types of gene rearrangement affect different VSG genes expressed within one sequentially derived series of trypanosome clones. It is evident that the 'expression-linked copy' model for VSG gene expression, as proposed by workers who observed extra copies of VSG genes in expressed states[3,4,12], does not describe the expression of the VSGs B and C in this series of trypanosome clones. Therefore, there must be some other mechanism controlling expression of these genes.

VSGs B and C are frequently expressed in relapse populations in animals infected with other ILTAR 1 clones (J. J. Doyle, personal communication), whereas VSG A is only rarely observed in relapse populations. This difference in VSG expression may be related to the different types of nuclear DNA rearrangements described here.

We thank Dr J. J. Doyle for providing trypanosome clones, Dr G. Matthyssens for critical reading of the manuscript and Ms Catherine Munyua for typing.

Received 2 February; accepted 21 April 1982.

1. Gray, A. R. & Luckins, A. G. in *Biology of the Kinetoplastida* Vol. 1 (eds Lumsden, W. H. R. & Evans, D. A.) 493–552 (Academic, London, 1976).
2. Williams, R. O., Young, J. R. & Majiwa, P. A. O. *Nature* **282**, 847–849 (1979).
3. Hoeijmakers, J. H. J., Frasch, A. C. C., Bernards, A., Borst, P. & Cross, G. A. M. *Nature* **284**, 78–80 (1980).
4. Pays, E., Van Meirvenne, N., Le Ray, D. & Steinert, M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2673–2677 (1981).
5. Williams, R. O., Young, J. R., Majiwa, P. A. O., Doyle, J. J. & Shapiro, S. Z. *Am. J. trop. Med. Hyg.* **29**(5), 1037–1043 (1980).
6. Williams, R. O., Young, J. R., Majiwa, P. A. O., Doyle, J. J. & Shapiro, S. Z. *Cold Spring Harb. Symp. quant. Biol.* **45**, 945–949 (1981).
7. Doyle, J. J. in *Immunity to Blood Parasites in Animals and Man* (eds Miller, L., Pino, J. & McKelvy, J. J.) 27–63 (Plenum, New York, 1977).
8. Young, J. R., Donelson, J. E., Majiwa, P. A. O., Shapiro, S. Z. & Williams, R. O. *Nucleic Acids Res.* **10**, 803–819 (1982).
9. Rice-Ficht, A. C., Chen, K. K. & Donelson, J. E. *Nature* **294**, 54–57 (1981).
10. Hoeijmakers, J. H. J., Borst, P., van Den Burg, J., Weissman, C. & Cross, G. A. M. *Gene* **8**, 391–417 (1980).
11. Thomas, P. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201–5205 (1980).
12. Bernards, A. *et al. Cell* **27**, 497–505 (1981).
13. Rice-Fitch, A. C., Chen, K. K. & Donelson, J. E. *Nature* (in the press).
14. Wahl, G. M., Stern, M. & Stark, G. R. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3683–3687 (1979).
15. Rigby, P. W. J., Dickman, M., Rhodes, C. & Berg, P. J. *J. molec. Biol.* **113**, 237–251 (1977).
16. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564 (1977).

Post-translational modification of tubulin dependent on organelle assembly

Karen J. Brunke, Philip S. Collis & Donald P. Weeks

The Institute for Cancer Research, The Fox Chase Center, Philadelphia, Pennsylvania 19111, USA

Microtubules are involved in a wide variety of cellular functions[1–3] and are major components of many subcellular structures (for example, the centriole, mitotic spindle, cytoskeleton and flagellar apparatus). The ability of microtubules to serve in such a diversity of functions may be accounted for, at least in part, by heterogeneity in the constituent types of α- and β-tubulin subunits[4–10] in different microtubules. Higher eukaryotes generally possess several tubulin genes and this may account for part of the heterogeneity of tubulin subunits[11–13]. However, the unicellular eukaryote, *Chlamydomonas reinhardii*, has only two α- and two β-tubulin genes[14–16], which suggests that some of the variation in tubulin subunits seen in this organism[5,8] may arise as a result of post-translational modification. This notion is supported by the experiments of Lefebvre *et al.*[8] who compared flagellar tubulins with those produced by *in vitro* translation of tubulin mRNAs. Here we show that a form of α-tubulin subunit apparently confined to the *Chlamydomonas* cell body is converted, post-translationally, into a flagellar form of α-tubulin and that this modification is dependent on flagellar assembly. We discuss possible mechanisms for these modifications and their implications for the generation of unique types of microtubules having specialized functions within the cell.

0028-0836/82/230516—03$01.00

When flagella are removed from *Chlamydomonas* gametes, the regeneration of a new set of flagella begins after ~10–12 min[17]. The initial rate of outgrowth is quite rapid[17] (~0.5 μm min^{-1}) but diminishes exponentially as the flagella lengthen. The new flagella approach their final length by ~90 min. During flagellar outgrowth, the synthesis of tubulin, the major structural protein of the flagellum, is rapidly induced[14,18,19]. Our studies have demonstrated that the synthesis of α- and β-tubulin subunits begins in less than 10 min after deflagellation and then rapidly increases to maximal rates in ~30–40 min. The autoradiograph shown in Fig. 1 illustrates this; it also reveals that in the SDS-polyacrylamide gels chosen for these experiments, the α-tubulin subunits produced from 10 to 90 min after flagellar excision can be resolved into two discrete bands. We have designated the upper band as α_f as it co-migrates with the predominant species of α-tubulin in the flagellum (see track 8 and also Fig. 2). The lower band co-migrates with the α-tubulin subunit species found exclusively in the cell body (see Fig. 2). Production of the α_f tubulin subunit is detectable neither during the first 10 min after flagellar removal nor beyond the time (~90 min) when flagellar outgrowth has been completed. The fact that α_f does not accumulate in the body of the cell and is produced exclusively during the period of flagellar assembly suggests that a portion of the newly synthesized α-tubulin subunits is converted to a modified form in a process associated with flagellar outgrowth. The results of the following experiments support this conclusion.

If mRNA is isolated from cells 45 min after deflagellation, when the production of α_f is at its peak (Fig. 1), and then translated in the wheat germ *in vitro* translation system[18], an α-tubulin polypeptide is produced which co-migrates with the α-tubulin of the cell body (Fig. 2a, lanes 4 and 7). There is no detectable synthesis of a polypeptide having the same electrophoretic properties as α_f (Fig. 2, lane 4 compared with lanes 3 and 5). These results suggest that either a mRNA encoding α_f exists but is not translatable *in vitro* or all cellular α-tubulin mRNAs are translated faithfully *in vitro* and produce α-tubulin which is identical (or almost identical) to the α-tubulin synthesized and stored in the cell body. If the latter is the case, then α_f could only arise as a result of post-translational modification. This interpretation is consistent with that of Lefebvre *et al.*[8] who performed similar *in vitro* experiments using the rabbit reticulocyte translation system. Nevertheless, the possibility still exists that a putative α_f mRNA does exist and that neither the wheat germ translation system nor the reticulocyte *in vitro* system is capable of translating it into a mature polypeptide. To establish that α_f is indeed the product of post-translational modification, we performed two additional sets of experiments.

In the first, we took advantage of the fact that tubulin induction following deflagellation is not dependent on subsequent flagellar outgrowth[17]. Flagellar regeneration was blocked in this experiment by treating deflagellated cells with colchicine. A comparison of tubulin species produced by deflagellated cells in the absence (Fig. 2*b*, lane 1) and presence (lane 2) of colchicine revealed that the α_f subunit was produced by the non-treated cells but not by a sample of the same cells treated with colchicine immediately after flagellar excision. As both samples were labelled for the same time period after deflagellation, these results also indicate that the appearance of α_f is dependent on flagellar outgrowth. (To ensure that the load size of protein in any of the samples did not influence the relative migrations of α- and α_f-tubulin subunits, radioactive proteins from flagella or cell bodies were combined with appropriate samples of non-radioactive proteins from the same sources as described in Fig. 2 legend, and run in Fig. 2*b*, lanes 3, 6 and 7.)

In the second experiment, cells were deflagellated and allowed to produce tubulin subunits for 120 min (Fig. 2*c*, lane 1) during which time flagella were regenerated (lane 4) and reserves of radioactively labelled tubulin accumulated in the cytoplasm (lane 6). At 120 min, cycloheximide was added at a concentration of 10 μg ml^{-1} which blocks essentially all protein synthesis on 80S ribosomes in *Chlamydomonas* (ref. 17; Fig. 2*c*, lane 3). The cells were then deflagellated[20], rapidly chilled, and centrifuged to separate out the labelled flagella. The pelleted cell bodies, which were free of contaminating flagella as judged by microscopic examination and by the lack of α_f-tubulin subunits (lane 6), were resuspended in fresh medium and allowed to regenerate flagella for 1 h. As expected[18,21], the cytoplasmic reserves of flagellar proteins allowed the regeneration of flagella that were on average 4–5 μm long (compared with 12–15 μm for full-length flagella). These short flagella were removed and separated from the cell bodies. Examination of the α-tubulin in the flagella regenerated in the presence of cycloheximide revealed that α_f-tubulin subunits were produced even in the absence of new tubulin synthesis (Fig. 2*c*, lane 5). None of the radioactively labelled α-tubulin that was converted to α_f during regeneration in the presence of cycloheximide was found in the body of the cell; α_f was found only in the flagella.

Taken together, these data provide strong evidence that the major α-tubulin subunit of the flagellum is the product of post-translational modification and, more importantly, that this modification is coupled to flagellar assembly. Two mechanisms can be envisaged to account for the close coupling of post-translational modification of α-tubulin and flagellar outgrowth. In the first, α-tubulin is modified in the cytoplasm only during periods of flagellar elongation and, once modified, is rapidly and quantitatively transported into the growing flagellum. The second, and perhaps more attractive hypothesis, is that

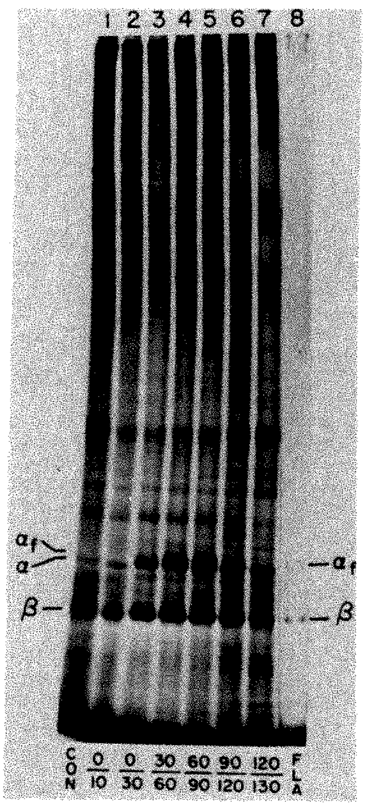


Fig. 1 Tubulin induction following flagellar excision in *C. reinhardii*. Autoradiograph of a SDS-polyacrylamide gel containing $H_2{}^{35}SO_4$-labelled proteins produced at various times before and after flagellar excision. Preparation of gametic cells by transfer of vegetative cells to nitrogen-deficient media and deflagellation of these cells by *p*H shock were performed as described previously[17]. Gametic cells were pulsed with $H_2{}^{35}SO_4$ before deflagellation (CON) and for the 10- or 30-min periods after flagellar removal shown below each track (for example, 30/60 denotes labelling from 30 to 60 min after deflagellation). Labelling conditions were as previously described[17] and included the use of chloramphenicol (250 μg ml^{-1}) to prevent the synthesis of chloroplast proteins. Samples having ~2×10^6 c.p.m. were loaded on to each track of the 15 cm-long 8% polyacrylamide gel and the proteins separated by their differential electrophoretic mobilities[18]. α Designates the α-tubulin subunit which migrates as a protein with an apparent molecular weight of 56,000; α^f, the predominant α-tubulin subunit of the flagellum; β, the β-tubulin subunit of molecular weight 53,000. Track 8 contains radioactively labelled flagellar markers (2×10^5 c.p.m.).

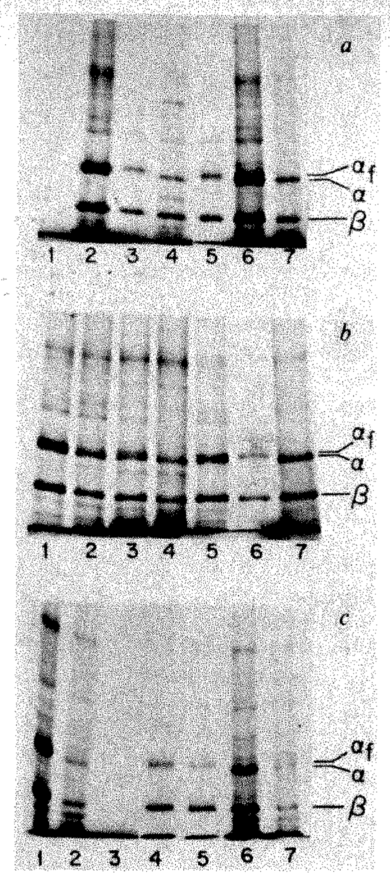


Fig. 2 Analysis of α- and α_f-tubulin subunit production. Autoradiographs of 8% polyacrylamide gels (containing 0.1% SDS) in which samples from cells and flagella labelled *in vivo* and samples of *in vitro* translation products ($\sim$1–2 $\times 10^5$ c.p.m. per track) have been separated by electrophoresis. *a*, Comparison of tubulin subunits from deflagellated cells labelled *in vivo* with ^{35}S-methionine-labelled products from the *in vitro* translation of *Chlamydomonas* mRNA. Procedures for the isolation of polyadenylated mRNA from deflagellated cells and its translation *in vitro* in the wheat germ translation system were as previously described[18]. *In vitro* translation reactions were performed in the absence (track 1) or presence (track 4) of mRNA isolated from fully-induced deflagellated cells. For comparison of these *in vitro* translation products with radioactive proteins produced *in vivo*, tracks 2 and 6 contain samples of cells that were deflagellated and labelled for 60 min with $^{35}S\text{-}H_2SO_4$. An aliquot of these same cells was deflagellated and flagella and cell bodies separated by density gradient centrifugation[20] to allow further comparison of flagellar tubulin (tracks 3 and 5) with tubulin contained in the body of the cell (track 7). The electrophoretic migration of the tubulin subunits produced *in vitro* and *in vivo* was not influenced by differences in protein load sizes as demonstrated by the following combinations of components: track 2, a mixture of the *in vitro* translation components of track 1 with proteins from deflagellated cells labelled *in vivo*; track 3, a mixture of the *in vitro* translation components of track 1 and flagella labelled *in vivo*; 5, unlabelled cell bodies and labelled flagella; 7, unlabelled flagella and labelled cell bodies. *b*, Effects of colchicine on the production of α- and α_f-tubulin subunits by deflagellated cells. Gametic cells were deflagellated and labelled with $^{35}S\text{-}H_2SO_4$ for 60 min in the absence (tracks 1 and 5) or presence (track 2) of colchicine at a concentration (1.5 mg ml^{-1}) which prevents flagellar regeneration. Control and marker samples were as follows: 4, proteins from labelled, non-deflagellated cells; 3, labelled non-deflagellated cells plus labelled flagella; 6, labelled flagella plus unlabelled cell bodies; 7, labelled cell bodies plus unlabelled flagella. *c*, Conversion of pre-labelled α-tubulin subunits into α_f-subunits during flagellar regeneration in the presence of cycloheximide, a protein synthesis inhibitor. Gametic cells were deflagellated and labelled with $^{35}S\text{-}H_2SO_4$ for 120 min (track 1). Non-deflagellated cells were similarly labelled in the absence (track 2) or presence (track 3) of cycloheximide (10 μg ml^{-1}). At the end of the labelling period, cycloheximide was added and a sample of the cells which had regenerated flagella were again deflagellated in order to allow separation of labelled flagella (track 4) from labelled cell bodies (track 6) by density gradient centrifugation. The pelleted cell bodies were resuspended in fresh medium containing cycloheximide and allowed to regenerate new flagella for 60 min in the absence of protein synthesis, then the cells were once more deflagellated to separate flagella (track 5) and cell bodies (track 7). In this case, after the cell bodies and flagella had been separated by centrifugation, the flagellar fraction was centrifuged at 30,000 r.p.m. for 45 min in a Beckman SW 40.1 rotor (as opposed to the usual centrifugation at 16,000 r.p.m. for 30 min in a Sorvall SS 34 rotor) to ensure the complete recovery of the short flagella regenerated in the presence of cycloheximide.

unmodified tubulin subunits move into the flagellum and are modified at or near the site of active microtubule assembly. In the latter case, localization of the enzymes responsible for tubulin modification might be accomplished by their binding to structures closely associated with the growing tips of the axonemal microtubules (for example, the distal filaments described by Dentler[22]).

The present and earlier[6,23] observations of flagellar-specific tubulins suggest that, in general, the ability of microtubules to perform a specific function in a particular subcellular structure may be influenced by the types of tubulin subunits they contain. Differences in the tubulin composition of microtubules can, in turn, be influenced by either genetic variability between different tubulin genes in the cell or post-translational modification of tubulin subunits. The presence of multiple tubulin genes in several higher organisms[11,12] indicates at least the potential to generate subunit diversity at the genetic level. However, two observations raise some question as to the role of genetic variability in the production of different classes of microtubules within a single cell. First, in *Aspergillus*, a heat-sensitive mutation in a single β-tubulin gene[24] seems to have pleiotropic effects on several microtubule-mediated functions—suggesting that tubulin produced from a single gene can be incorporated into most, if not all, of the microtubules of the cell. Secondly, in *Drosophila*, a mutation has been found in the structural gene for a β-tubulin that is expressed only in testis tissue[7]. The variant tubulin produced from this 'specialized' tubulin gene is apparently not confined to a particular type of microtubule, but rather has wide ranging detrimental effects on various microtubule-containing structures within the cell. Therefore, if the existence of organelle-specific tubulin subtypes is widespread, the production of these subtypes may well have to depend on the kinds of post-translational modifications reported here for the α_f-tubulin subunit of the *Chlamydomonas* flagellum.

Finally, one hypothesis that may be drawn from our results is that the enzymes required for flagellar α-tubulin modification may be located close to the site of axonemal microtubule assembly. In a broader context, this hypothesis suggests that the enzymes which modify tubulin subunits and thereby tailor microtubules to perform the specialized functions of a particular organelle may not be scattered throughout the cytoplasm but may be in close and perhaps exclusive association with the microtubules of that organelle.

These studies were supported by NSF grant PCM-80-03438, NIH grant CA-06927 and RR-05539, and by an appropriation from the Commonwealth of Pennsylvania. K.J.B. is a recipient of funds from the NIH (CA-09035).

Received 15 February; accepted 2 April 1982.

1. Kirschner, M. W. *Int. Rev. Cytol.* **54**, 1–71 (1978).
2. Otto, A. M. *Cell Biol. Int. Rep.* **6**, 1–18 (1982).
3. Huang, B., Ramanis, Z. & Luck, D. J. L. *Cell* **28**, 115–124 (1982).
4. Bibring, T., Baxandall, J., Denslow, S. & Walker, B. *J. Cell Biol.* **69**, 301–312 (1976).
5. Piperno, G. & Luck, D. *J. biol. Chem.* **251**, 2161–2167 (1967).
6. Stephens, R. E. *Biochemistry* **17**, 2882–2891 (1978).
7. Kemphues, K. J., Raff, R. A., Kaufman, T. C. & Raff, E. C. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3991–3995 (1979).
8. Lefebvre, P. A., Silflow, C. D., Wieben, E. D. & Rosenbaum, J. L. *Cell* **20**, 469–477 (1980).
9. George, H. J., Misra, L., Field, D. J. & Lee, J. C. *Biochemistry* **20**, 2402–2409 (1981).
10. Sheir-Neiss, G., Lai, M. H. & Morris, N. R. *Cell* **15**, 639–647 (1978).
11. Cleveland, D. W. *et al. Cell* **20**, 95–105 (1980).
12. Cowan, N. J., Wilde, C. D., Chow, L. T. & Wefald, F. C. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4877–4881 (1981).
13. Sanchez, F., Natzle, J. E., Cleveland, D. W., Kirshner, M. W. & McCarthy, B. J. *Cell* **22**, 845–854 (1980).
14. Minami, S. A., Collis, P. S., Young, E. E. & Weeks, D. P. *Cell* **24**, 89–95 (1981).
15. Brunke, K. J., Young, E. E., Buchbinder, B. U. & Weeks, D. P. *Nucleic Acids Res.* **10**, 1295–1310 (1982).
16. Silflow, C. D. & Rosenbaum, J. L. *Cell* **24**, 81–88 (1981).
17. Weeks, D. P., Collis, P. S. & Gealt, M. A. *Nature* **268**, 667–668 (1977).
18. Weeks, D. P. & Collis, P. S. *Cell* **9**, 15–27 (1976).
19. Lefebvre, P. A., Nordstrom, S. A., Moulder, J. E. & Rosenbaum, J. L. *J. Cell. Biol.* **78**, 8–27 (1978).
20. Witman, G. B., Carlson, K., Berliner, J. & Rosenbaum, J. L. *J. Cell Biol.* **54**, 507–539 (1972).
21. Rosenbaum, J. L., Moulder, J. E. & Ringo D. L. *J. Cell Biol.* **41**, 600–619 (1969).
22. Dentler, W. L. *J. cell. Sci.* **42**, 207–220 (1980).
23. Fulton, C. & Simpson, P. A. in *Cell Motility* (eds Goldman, R., Pollard, T. & Rosenbaum, J.) 987–1005 (Cold Spring Harbor Laboratory, New York, 1976).
24. Oakley, B. R. & Morris, N. R. *Cell* **24**, 837–845 (1981).

BOOK REVIEWS

A new nuclear holocaust

John Maddox

IN 1962, Mrs Rachel Carson discovered the harmful effects of pesticides in the pages of the *New Yorker*, and life has not been the same since for the chemical manufacturers. Earlier this year, Mr Jonathan Schell used the same vehicle to announce his personal discovery that nuclear weapons also have harmful effects, and in the process helped to fuel the influential campaign in the United States for a "nuclear freeze". Like *Silent Spring, The Fate of the Earth* is a republished collection of the magazine articles.

The result is compulsive reading but not necessarily a compelling book. The first half is the simpler and the more arresting, written as it is with restrained passion, suppressed rage perhaps. Yet Schell is not an out-and-out scaremonger but rather a cautious one. His "nuclear predicament" is his way of referring to people's indifference to the threat embodied in the nuclear arsenals — "the world has declined, on the whole, to think about them very much". Properly and persistently he reminds his readers that uncertainties must abound in attempts such as his to show what would happen in an all-out nuclear attack on the United States. Would more people be killed by prompt radiation from an airburst weapon than from fire or blast? Only time, says Schell, could tell.

But how can it be possible to write vividly about a topic, the destruction that might be caused by a nuclear attack, when thousands on thousands of words have already been cast in type? Schell succeeds for two reasons, one less creditable than the other.

The meaner trick is the assumption of discovery:

> We have lived in the shadow of nuclear arms for thirty years, so it does not seem too soon for us to familiarize ourselves with them . . . .

or

> a description of a full-scale holocaust seems to be made necessary by the simple but basic rule that in order to discuss something one should first know what it is.

There follows a list of the technical reports that Schell has referred to, but there is no acknowledgement of other writers in the same genre — Robert Junk with his *Brighter than a Thousand Suns,* for example. One's first reaction is resentment: why should Szilard, for example, rate a mention in Schell's account only as one "who called on a number of

The Fate of the Earth. By Jonathan Schell. Pp.244. UK hbk ISBN 0-224-02064-1; UK pbk ISBN 0-330-26915-1; US ISBN 0-394-52559-0. (Jonathan Cape/ Picador/Knopf: 1982.) Hbk £7.95, $11.95; pbk £1.95.

colleagues to keep the discovery [of fission, not then of the bomb] secret from the Germans"? Was he not the Manhattan Project man who wore his conscience most conspicuously on his sleeve? But charitably, one quickly comes to terms with Schell's assumption. To discover the prospective awfulness of the holocaust is more likely to catch the attention than to recount what has been told already. Schell is in any case properly diffident about his knowledge of distant historical events (using the phrase "it is reported" to describe what happened during the Cuban missile crisis of 1962); and then, perhaps, each generation that lives with the bomb needs its own prophet.

Schell's second trick, his literary skill, gives him a good chance of being that. His account of what the holocaust might be like *is* original, almost fresh. Fission, as fusion, is the exception that proves Einstein's rule that mass can be converted into energy. The novelty in fission is that the strong nuclear force was for the first time manipulated by people. One gram of uranium destroyed Hiroshima, but the bomb weighed four tonnes.

The literary rehearsal of the holocaust takes off effortlessly from this point. Schell says that in "this unique endeavour", "foresight is asked to perform a task usually reserved for hindsight". He is only a reluctant futurologist. He takes comfort from the circular slide-rule distributed with copies of Glasstone and Dolan's book *The Effects of Nuclear Weapons* to ground his calculations in someone else's reality, and he does his best to follow the *New Yorker's* precept that the whole world has an absorbing interest in what happens in Manhattan by exploding his mercifully hypothetical one-megaton weapon above the Empire State Building; all the buildings between Battery Park and 125th Street would be knocked down, the fireball would set the city alight. "From Greenwich Village up to Central Park, the heat would be great enough to melt metal and glass."

Schell, of course, is telling the truth. He is probably also right in his guess of what would happen if the existing stock of Soviet nuclear warheads were fired at a representative set of targets throughout the United States. Communications systems would be put out of action. Most cities not flattened would be burning. The few survivors would be either beastly or indifferent to each other, and hampered in their efforts to restart some kind of primitive society by their load of genetic defects and consequent somatic transformation. The critical question, for Schell, is whether the result would be to extinguish the human species or to make life on the Earth impossible, perhaps by the removal of the ozone cover, for all except a few plants and radiation-resistant insects. He acknowledges that the outcome is uncertain, but although there is "all the difference in the world" between the chance and the certainty of extinction, "morally they are the same".

Peroratively, half-way through his book, Schell remarks:

> In weighing the fate of the earth and, with it, our own fate, we stand before a mystery, and in tampering with the earth we tamper with a mystery. We are in deep ignorance. Our ignorance should dispose us to wonder, our wonder should make us humble, our humility should inspire us to reverence and caution, and our reverence and caution should lead us to act without delay to withdraw the threat we now pose to the earth and to ourselves.

This purple passage is the climax to the book. Tract though it may be, it is a well-intentioned tract. Its naivety can be forgiven as innocence. Its neglect of the agonies that have exercised people of all kinds since 1938 is less easily forgiven, but Mr Schell has an important point to make.

At this high point in his book, with only half the pages turned, readers are bound to wonder what can come next. An account, probably ironical, of the attempts that people have been making over the years to negotiate arms control agreements? A discussion of the danger of nuclear proliferation, and of the wayward policies of the nuclear suppliers in this regard? The answer, it turns out, is nothing like that but a chapter called "The Second Death", a kind of essay on the theme that even in strictly personal terms there is something worse than being killed — the whole species might be eliminated.

Nobody will quarrel with the conclusion. Did not even St Augustine, for whom the objective of every human being was to arrive safely in Heaven, acknowledge that even quite ordinary people could be fixated on the terrestrial future by their regard for their children and their children's children?

But Schell skates on thin ice in getting there.

Thus he asserts that scientific knowledge is different from other knowledge in that it cannot be forgotten. Really? And that

> a disturbing corollary of the scientists' inability even to foresee the path of science, to say nothing of determining it, is that while science is without doubt the most powerful revolutionary force in the world, no-one directs that force.

But Schell's real objective is to contrast and compare the notions of death and extinction, one personal, the other for the species. He rightly and evocatively finds a parallel between the extinction of wagon-loads of people in the concentration camps and the feared extinction of whole species. It is an ethical goal, he says, to seek to avoid extinction. Yet living as we do on top of a nuclear stockpile, we are "living with a lie".

The core of Schell's final essay is a comment on the doctrine of deterrence, a popular target these days. If the objective is to build a weapons system so powerful that the destruction of an adversary can be assured, how can success be made convincing without a demonstration? But detente is no good either: look at the way the Soviets made "crimes against detente" punishable, and how President Nixon pleaded exemption from the laws of the United States for the sake of a peaceful world. The real bugbear, Schell says, is sovereignty (but his book was written before the Falkland Islands conflict).

> . . . just as those who favour the deterrence policy . . . must in all honesty admit that their scheme contemplates the extinction of man in the name of protecting national sovereignty, so must those who favour complete nuclear and conventional disarmament, as I do, admit that their recommendation is inconsistent with national sovereignty . . . .

Read as a piece (and it does not take long), Mr Schell's book is thus an extraordinary let-down. "I have not sought to define a political solution to the nuclear predicament" but "I have left to others those awesome urgent tasks . . .". There's generosity for you; some other author, another four-part series in the *New Yorker*, perhaps?

What Mr Schell has forgotten (apart from the precursors whom he fails to mention) is that the simplest explanation of his opening conundrum — why are we mostly so indifferent? — is not a failure to imagine what nuclear war would do to us but a failure to devise political procedures for its sure avoidance. People have seen the difficulties and have sometimes lost heart. Mr Schell's recipe, if taken seriously, can only further depress them, for in his eloquent way he is saying that the problem would be manageable if only the world were an entirely different place. For a book that makes so much of people's ethical responsibilities, that is not merely a disappointing conclusion but an irresponsibly airy recipe for conduct. □

John Maddox is Editor of Nature.

Journals' review issue 1982

On October 7th *Nature* will publish a second review supplement devoted to science journals. Last year's supplement, covering journals first published between January 1978 and May 1980, appeared in *Nature* **293**, 341-369; see p.343 of that issue for details.

Criteria for inclusion of a journal in the 1982 issue are that:

- the first number appeared, or the journal was re-titled, between June 1980 and May 1981;
- it is published at least three times a year;
- the main language used is English.

Broadly, periodicals of professional interest to scientists will be considered for review, with the exception of abstracts' journals.

In addition it is hoped to cover publicly available newsletters, first published between January 1978 and May 1981, in that issue.

Publishers and societies are invited to submit four sample issues of periodicals satisfying the above criteria, ***including the first and most recent numbers***, to the Review Editor, *Nature*, 4 Little Essex St, London WC2R 3LF, England (London 836 6633 ext 2570) as soon as possible.

Affinity for industry and biomedicine

P.G.H. Byfield

Affinity Chromatography and Related Techniques: Theoretical Aspects/ Industrial and Biomedical Applications. Analytical Chemistry Symposia Series, Volume 9. Edited by T.C.J. Gribnau, J. Visser and R.J.F. Nivard. Pp.584. ISBN 0-444-42031-2. (Elsevier Scientific: 1982.) Dfl.195, $83.

AFFINITY chromatography has marched out of the research laboratory and into the world. That was to be the message of the symposium held in June of last year at Veldhoven, of which this book is a record, and the organizers emphasized it by bringing together similar numbers of participants from industry and academia. Meetings based on the discussion of a technique run the risk of being boring since the wide range of possible applications may contain only little of interest to an individual. The organizers overcame this by concentrating the contributions around biomedical applications and relevant industrial processes. This theme penetrated also into the contributions on theoretical considerations and developments in supports and chemistry; many of these were based on biomedical or industrial needs.

By far the most exciting current application of affinity chromatography is indeed to the biomedical field. Even the ability to simplify the isolation of useful compounds from biological sources will open up new possibilities in patient treatment and management. Several contributions describe the commercial production of proteins from human plasma by this technique, others indicate the way to future products, while one author laments that only lack of knowledge of their function inhibits the isolation of more plasma constituents — perhaps the physiologists will take up this challenge. Exploiting the high specificity of antibodies in therapy, as reagents and as affinity ligands, has often been hindered by difficulties in their isolation due to the high interaction energies between antigen and antibody. The successes described here in the fractionation of antisera and in the purification of vaccines will help many, particularly when combined with the practical conclusions in the theoretical paper by van Oss.

Some disquiet was expressed at the meeting that ligand leakage might contaminate products intended for clinical use with hazardous compounds, especially where organic dyes are the ligands. These problems should be easy to overcome. After all, what is affinity chromatography about if not to remove small quantities of a substance from large quantities of other material?

One forward-looking section of three papers considered the use of affinity systems directly in therapy, for the delivery of drugs to specific sites *in vivo* and for the extracorporeal treatment of blood to remove toxins, drugs and so on. Although there is a long way to go, these approaches to more specific treatment of disease are surely to be encouraged.

It is rare to read of an affinity chromatography system that is not based on cyanogen bromide-activated agarose, yet most meetings include sections on new supports and coupling methods. Why is this? One answer was evident at Veldhoven. Both large-scale work and the new high-pressure liquid affinity chromatography systems need robust supports such as the glass, silica and polymers described by several speakers.

Many good posters were shown at the meeting and it is a pity that, tantalizingly, only the titles, not their abstracts, have been published in the book. Nonetheless those interested in using the technique rather than in studying the phenomenon will find much of value in this volume; even the theoretical section contains advice which will be of much practical use. □

Peter Byfield is a Scientific Officer in the Endocrinology Research Group at the Clinical Research Centre, Harrow, Middlesex.

Uneven depths in marine geology

Philip Weaver & Lindsay Parson

The Oceanic Lithosphere. Edited by Cesare Emiliani. Pp.1,738. ISBN 0-471-02870-3. (Wiley: 1982). £110, $195.

SINCE the inception of the Deep Sea Drilling Project, which has played a major role in providing data from previously unsampled horizons in both sediments and the basaltic basement, marine geology has emerged from infancy and into an exciting period of discovery. More recent work using manned submersibles has proved invaluable for the precision sampling of specific areas of interest. Investigations have not been restricted to these data, however, and bottom sampling and geophysical work outside the DSDP are providing the basis for new studies, such as the extensive use of gravity and piston core sampling in the investigation of Quaternary sediments (for example in the CLIMAP project).

The Oceanic Lithosphere, the seventh volume of *The Sea* series, attempts to bring together some of the new discoveries and to provide reviews of the major fields of research. Its scope is therefore enormous, and no one book could hope to cover all aspects fully. The subject matter of the 35 chapters falls naturally into two sections, the first dealing with upper mantle and crust, the second with the sedimentary record.

It is noteworthy that a number of the review articles — such as those on heat probe and on magnetics, and on the composition and evolution of the ocean crust — constitute works of equal, if not greater value to researchers of a general geological or geophysical bent seeking broad appraisals of the ocean basins than to the specialist oceanographers to whom this book claims to be directed. In contrast, or perhaps as a balance, a number of detailed discussions — such as those considering helium isotope work and radiogenic gas retention — will be of far more limited appeal, and may have been included at the expense of swelling an already impressively over-large tome.

Nevertheless, valuable contributions are made in the form of studies of ophiolite suites and their role in the understanding of the evolution of present-day oceanic crust. Petrological analyses of oceanic basement are well represented, and in particular the review by Fox and Stroup provides a comprehensive synthesis of work on oceanic crust with its discussions of composition, seismic structure, heat flow results and magnetism.

The section on marine sediments covers a range of topics from sediment composition and origin to the evolution of ocean basins. Apart from red clays, most oceanic sediments are biogenic in nature, and analysis of their microfaunas enables us to understand past climates and ocean circulation patterns, and to identify changes in the normal sedimentary regime such as hiatuses, anoxic intervals and periods of mass extinctions. *The Oceanic Lithosphere* necessarily devotes several chapters to reviews of the various microfossil groups, providing comprehensive summaries of the particular group, their present-day and past distributions, and their value to oceanography. These are followed by chapters reviewing the deep-sea record and the dynamics of the Earth's sediment system as a whole.

One of the areas of particular interest at the moment is the study of sedimentary processes. Identification of sediment drifts, submarine slides and erosional channels, and the mapping of turbidite flows shows us that the ocean floor is not necessarily stable. Regrettably, this area of research is not covered in the book and readers seeking an overview of the subject could be left with the misconception that the oceanic sedimentary record consists of a layered sequence of undisturbed sediments.

Taken as a whole the volume offers a somewhat patchy review of the oceanic lithosphere. It would have been more logical to produce it as two separate parts as with other volumes in this series, and perhaps to add a few more topics to each part. Largely as a result of the amalgamation of these two parts, *The Oceanic Lithosphere* appears destined for the reference collection kept on a sturdy shelf. □

Philip Weaver and Lindsay Parson are at the Institute of Oceanographic Sciences, Wormley, Surrey.

All round the world of algae

Brian Whitton

The Ecology of Algae. By F.E. Round. Pp.653. ISBN 0-521-22583-3. (Cambridge University Press: 1981.) £60, $130.

PROFESSOR Round has presented us with the type of book which is becoming rarer and rarer, a major monograph written by just one person. Algae are a decidedly heterogeneous group of organisms, so the author has faced up to a challenge which would daunt most editors of multi-author symposium proceedings. *The Ecology of Algae* has well over two thousand references integrated into the text, so it approaches in scope such famous works as F.E. Fritsch's *Structure and Reproduction of the Algae* (Cambridge University Press; 1935, 1945) or one of the volumes of G.E. Hutchinson's *A Treatise in Limnology,* published by Wiley. The question, then, is just how well has the author succeeded in covering this vast topic.

Professor Round has himself published papers on freshwater and marine environments, and has travelled and researched in many countries. The book obviously benefits from this, as both types of environment are given equally sympathetic treatment and examples are chosen on a world-wide basis. There is only one specifically marine chapter, on coral reefs. Throughout the rest of the book the algal ecology of freshwater, subaerial and marine environments is brought together, usually in adjacent sections, but sometimes in a single integrated account. This approach is often very successful in showing similarities between different environments, particularly among the phytobenthos communities which were the subject of much of the author's earlier research. As far as I can judge, most topics are dealt with thoroughly and competently, but in comparison with the space given to the behaviour of algae, treatment of the environment is brief. Information on nutrient cycles is particularly sparse and rather uncritical.

The text steers a reasonable course between the summary of numerous case studies and synthesis; there is none of the nonsense associated with the glib generalizations of several modern algae textbooks. A great deal of information is packed into each section, so the text often requires hard concentration; many readers will probably use the book more as a means of finding their way into the literature than as one from which whole chapters are read in a single sitting. I enjoyed especially Round's own comments on the state of particular areas of research. Such asides can be decidedly annoying, but Round's often show considerable insight and are so sensible that I found them most helpful.

The book is well produced, with numerous figures, useful summaries at the end of each chapter, few printing errors and accurate references. An appendix summarizes the current state of algal systematics, though (to quibble) I would disagree with the inclusion of the Prochlorophyta as a phylum. The main text reviews a large body of literature with a sweep that few, if any, other phycologists in the world could attempt. Research on ecology dates much more slowly than that on physiology and biochemistry, so *The Ecology of Algae* is likely to remain an important text for at least the next decade. I hope librarians will find the money to buy this book, whatever the state of their finances. □

Brian Whitton is Reader in Botany at the University of Durham.

Ape reproduction: science and ethics

Alison Jolly

Reproductive Biology of the Great Apes: Comparative and Biomedical Perspectives. Edited by Charles E. Graham. Pp.437. ISBN 0-12-295020-8. (Academic: 1981.) $52.50, £34.80.

THIS is an unusual book: it achieves just what it sets out to do. It is a handbook that summarizes existing knowledge of ape reproduction, and in the process also raises new ideas which will shape future research. And yet, for those few people who read the entire volume, the clearest impression is not the zoological synthesis the various authors achieve, but an ethical chasm which divides them.

To be sure, the synthesis is impressive: studies of ape reproductive tracts, secretions, hormonal control, pregnancy and birth, endocrine development, sexual behaviour in laboratory or field, as well as conditions of caging and handling for different studies, are brought together and analysed thoroughly. The book shows that the field observations do correlate with laboratory studies of anatomy and physiology; one can indeed illuminate the other. This volume will thus be an essential reference source.

But how do these ape studies apply to human beings? Here we reach the gulf between the indoor and the outdoor scientists. Each of the physiological articles ends by saying that ape reproductive parameters are much closer to those of human beings than to rhesus monkeys'. Each then concludes that we must conserve the great apes for use in medical experiments unethical with human beings, though most add that both ethics and commercial good sense mean limiting damage to such "models" to a minimum.

The field workers, instead, want apes conserved for their own sake and for the scientific and aesthetic pleasure they give us. Perhaps they even share my distaste for calling a living animal a "model", which implies that it exists only as a human surrogate. They certainly emphasize conservation of wild habitat with its mothers that dote five years on one infant before bearing the next, instead of "rational" rearing in nurseries. Battery ape production means removing infants from the mother at birth, if destined for experimental use, so that mothers skip lactational amenorrhea and get on with the next pregnancy. At present, a breeding nucleus of infants must be left for two years with the mother before removal if they are to become emotionally capable of adult sexual intercourse. Later perfection of artificial insemination might remove any "need" for either ape sex or ape motherhood.

I exaggerate, but only slightly. Understanding ape reproduction is desperately important to maintain the captive populations: 1,000 chimpanzees in the United States; far fewer gorillas and orang utans. In Great Britain and Ireland, as of January 1980, zoos held 167 chimpanzees, 63 gorillas and 61 orang utans, few of them breeding adults. We may argue cynically that we shall not pay to support these creatures and their remaining wild cousins unless we selfishly think of them as "models", or we may frankly argue that they should be preserved for other reasons. Of the physiologists who have contributed to this book, only R.V. Short has the courage of his logic as well as his convictions to conclude:

> Humans not only are more available, but also infinitely more cooperative than apes; human reproductive biology is more likely to be pioneered in the human, and it is seldom that the basic facts will first be revealed in an ape, and subsequently extrapolated to man . . . If the only great apes to survive in the 21st century are to be found in zoological gardens and laboratories, we will have lost forever many of the vital behavioral clues which explain why these animals are built the way they are. The great lesson of evolution is that form reflects function. Bequeathed only a caged, or worse still, a stuffed gorilla, we would have no way of determining the adaptive significance of any of its anatomical features . . . if we can begin to correlate form with function in our closest relatives, then by analogy we can gain new insights into the selective forces that have made humans what they are today.

Alison Jolly is a Guest Investigator at The Rockefeller University, New York.

Knowing one's place

R. H. S. Carpenter

Human Visual Orientation. By Ian P. Howard. Pp.697. ISBN 0-471-27946-3. (Wiley: 1982.) £25, $59.95.

OUR senses provide us with two distinct kinds of information: *what,* and *where.* Neurophysiologists, beguiled by the complexities of *what,* have tended to neglect *where;* yet the location of stimuli is far from being the simple matter one might at first suppose.

Objects can only be localized relative to the organ that senses them; if this is itself moveable, then we need to know not just where the object is relative to the sense organ, but also where the sense organ is relative to the body. In the case of vision, we have a highly mobile eye mounted on a freely turning head, attached to a flexible body which can be moved about with the feet. Add to this the large number of degrees of freedom enjoyed by our arms and hands, and it can be seen that even an

action as simple as picking up a pencil demands the complex processing of information not just from our eyes, but from sense organs in our muscles, skin and joints, from the vestibular receptors that tell us about the motion and attitude of our head relative to gravity, and from the stored expectations that we have of the way the world is structured.

To a very large extent, this integration of various kinds of information must be learnt: unlike a surveyor's theodolite, our senses do not come to us ready-calibrated in degrees and minutes of arc, but must be scaled by our own brain through its experience of the way in which one kind of sensation tends to be correlated with another. If we wear prisms that deviate our sight to one side, though at first we reach out for visual objects at the wrong place, after only a few trials our brain adjusts itself to the new situation and no errors are made. This aspect of sensory location has stimulated a great deal of interest over the past few years, for it enables one to make attractively quantitative measurements of the learning processes involved. The advantage of studying sensorimotor coordination rather than perception is that we then work with a system with a clearly defined output as well as input; in the latter case there must always be some doubt as to the nature of whatever it is that is doing the perceiving.

Behind a deceptively modest title, *Human Visual Orientation* hides a wealth of detailed and remarkably up-to-date information about nearly all aspects of this very wide subject. Particularly welcome, in a field where slipshod terminology can easily lead to tangled confusion (how, for example, does one rigorously define the direction of gaze?), are the author's succinct definitions of technical terms, and his clear and intelligent reviews of salient facts in an area in which many important observations date back to the nineteenth century. Many topics which the title of the book might not lead one to expect — the neurophysiology of eye movements, for one — are summarized with equal clarity and expertise.

In other respects, however, the author has perhaps been too modest in his aims: his explicit omission of motion perception and other dynamic aspects of orientation, and perhaps especially his suspicion of model-building and his reluctance to synthesize and integrate on more than a moderate scale, may leave some readers unsatisfied. But this does not detract significantly from what is otherwise a first-rate book, a richly-veined mine of information that will undoubtedly become a standard work of reference and provide much to ponder upon for many years. □

Roger Carpenter is a Lecturer in Physiology at the University of Cambridge. He is author of Neurophysiology, *shortly to be published by Edward Arnold, and of* Movements of the Eyes *(Pion, 1977).*

To whom does the United States belong?

Lynton K. Caldwell

The Land Use Policy Debate in the United States. Edited by Judith I. de Neufville. Pp.269. ISBN 0-306-40718-3. (Plenum: 1981.) $29.50, £18.59.

TO UNDERSTAND the land use policy debate in the United States one must first recollect its historical context. During the colonial era, and for decades thereafter, settlers were drawn to the New World by the prospect of free land unburdened by the feudal legacy of taxes and servitudes. And whereas general English common law regarding ownership and tenancy of land was brought to North America, specific restrictions were left behind. Moreover the early years of colonization occurred at a time when English landowners were shaking off feudal exactions. In any event, controls by government were largely unenforceable on the frontiers of settlement. Then, as public authority and jurisdiction were extended to unappropriated lands, the guiding object of public policy was to move ownership as rapidly as possible into private hands.

For a century after independence, speculation in the purchase and sale of public lands was one of the principal roads to wealth in the United States. And the most respected figures in public life, not excepting George Washington himself, did not hesitate to engage in it. The privatization of land was carried further in North America than perhaps in any other country in the modern world. Attempts to impose even minimal controls on the uses of land have not come easily. In 1970 the Senate of the United States considered a National Land Use Planning Act, which was intended to encourage states and localities to adopt forms of democratic planning that would avoid at least the worst conflicts and demonstrable misuses of land. Endorsed by President Nixon, this measure passed the Senate on three occasions but never reached the floor of the House of Representatives, being blocked by a combination of land speculators, developers, farmers and ordinary citizens who feared that the government was about to take away the right to do as they pleased with the land that they owned.

Against this background, one may better understand the 20 essays that make up this book. They evolved out of a colloquium held in 1977 under the sponsorship of the Lincoln Institute of Land Policy at Cambridge, Massachusetts. In view of the long history of frustration in land use planning in the United States, it was appropriate that the focus of the colloquium was on latent values in land use policy. It should be apparent that a significant measure of change in popular understanding and attitude will have to occur before rational land use policy will become effective in the United States.

The scope of the colloquium, and the book, includes considerations of land use as affected by economic, social and environmental policy, the role of markets, the law and public participation. Finally, the authors raise the possibility of a new

The geologist's wife

TO HER HUSBAND SETTING OFF
UPON AN EXCURSION

Adieu then, my dear, to the Highlands you go,
Geology calls you, you must not say no;
Alone in your absence I cannot but mourn,
And yet it were selfish to wish your return.

No, come not until you have searched through the gneiss,
And marked all the smoothings produced by the ice;
O'er granite-filled chinks felt Huttonian joy,
And measured the parallel roads of Glenroy.

And if, in your wanderings, you chance to be led
To Ross-shire or Moray, to see the Old Red,
Oh still, as its mail-covered fishes you view,
Remember the colour is love's proper hue.

Such being your feelings, I'll care not although
You're gone from my side — for a fortnight or so;
But know, if much longer you leave me alone,
You may find, coming back you have two wives of stone!

These verses, four of the eight in *The Geologist's Wife* — written by the anonymous but evidently love-lorn "Susan" in 1847 — are taken from *The Poetry of Geology,* recently published by George Allen & Unwin. The anthology is edited by Robert M. Hazen and costs £4.95, $9.95.

land ethic and identify a series of problems relating to political control of land use. The editor of this volume, who is a member of the Department of City and Regional Planning of the University of California at Berkeley, pulls together the principal themes and viewpoints of the authors in a final essay.

Four general policy positions emerge in the discussions, although not all of the viewpoints are shared by all authors. The first is that the test of good land use policy is the procedure by which it was adopted. This opinion divides between those who emphasize participatory democracy and those who advocate market processes. The second position is that a policy is good if it truly represents the public interest. This view of "goodness" depends less upon what people think than upon ostensibly objective assessments of the implications — economic, ecological and ethical — of particular land use decisions. A third position is that a good land use policy is one that is allied with technological or problem-solving criteria. From this viewpoint land use capability, as assessed by technological and economic possibilities, is the significant factor in determining the land's best and highest use. Fourth, and final, is the notion that good land use policy must be "socially just", correcting, or at least not aggravating inequalities of wealth and power in society.

Although the viewpoints taken in this collection of essays are not new, they seem to require periodic restatement in a country in which deep-rooted emotional and psychological problems inhibit the development of any consistent or rational policy regarding land. Ann L. Strong, the most widely known of the authors, is surely right in her contention that in the United States there has always been an ultimate public interest in the care and custody of land. No less a libertarian than Thomas Jefferson made this explicit. But the myth of absolute and uninhibited private ownership has dominated public policy even to the demonstrable detriment of most private land owners.

The principal value of this collection of essays is that it reiterates the terms of the debate and keeps the discussion going at a time when the prospects for the reforms that most of the authors advocate appear unpromising. I agree with Ann Strong in the title of her essay, "Land As A Public Good: An Idea Whose Time Has Come Again", but I recall that the idea of a national land ethic implemented through public law has come among Americans before, and has not been received. I believe that the force of circumstances will inevitably cause the concept of land as a public good to be generally accepted, but I am not yet persuaded that that time is now. □

Lynton K. Caldwell is Bentley Professor of Political Science at Indiana University, and participated in drafting the United States National Land Use Policy Act of 1970.

A waste of time for everybody?

Susanna Millar

Animal Play Behavior. By Robert Fagen. Pp.684. Hbk ISBN 0-19-502760-4; pbk ISBN 0-19-502761-2. (Oxford University Press: 1981.) Hbk £21, $29.95; pbk £10.50.

THE cover notes and preface to this book declare that Fagen has broken new ground by adopting an evolutionary approach to play, and that he has solved an important biological paradox. This is a somewhat inflated claim. With the single exception of behaviourism, now long outmoded, there is no theory of play, and no developmental theory, that has not assumed an evolutionary context or has operated in a biological vacuum since Spencer first tried to account for play, and Darwin started child watching. The introduction explains that only analyses based on sociobiology and behavioural ecology count. Fagen anticipates that this will cause offence. It is not at all clear why, except that the aggressive, polemical style seems to invite it. Muted hope, that the proposals will justify the claims made for them, is a much more likely response.

Fagen does, nonetheless, make a worthwhile contribution; more to method than to theory, but potentially all the more useful for that. He proposes to apply formal modelling techniques to play, using assumptions based on current genetic and evolutionary ideas. Formal mathematical and computer models, designed to simulate behaviours of complex systems, can be used to investigate the properties of the systems and are widely used in research in biology, genetics and psychology. Fagen's examples of possible model fitting with questions related to play, given mainly in appendices, are much the most interesting parts of the book and provide clear pointers to what could be done.

The main problem about play, according to Fagen, is that animals geared to survival apparently "waste time" by playing. Actually, this is more a terminological quibble than a biological paradox. The "useful/useless" dichotomy depends on the system of classifying behaviour by the outcome, goal or end-state which is obvious to the observer. It is arbitrary and, therefore, provides no reason against analyses of play in terms of survival value; or against applying the theory of frequency-dependent natural selection of genetically based variations to play behaviours. But there is also a genuine difference between play and behaviours such as sex and aggression, which here gets rather lost. This is the paradoxical relation between play activities and the contexts (outcomes) in which they occur. It poses no problem for biological principles, and the "useful/useless" distinction is irrelevant to it. But the relation between activity and context does differentiate play from other behaviours, and it deserves rather more attention than it is accorded by Fagen.

By far the largest part of the book is devoted to the assumptions that Fagen has selected to explain play. These are given as chapter and section headings for his arguments. Since we do not as yet have the results of independent tests, these arguments carry no more weight than any other explanation based on observational and experimental studies. Fagen argues that play is an innate general tendency to prepare for later versatility, with possible connections to art and science. This view is not new. More importantly, it is not implied by evolutionary principles. The fact that it is, of course, unlikely that play behaviours would occur if they did not help with "fitness" and species survival in various ways does not by any means logically imply that all types of play must have the same function, or depend on a single mechanism. As Fagen is aware, such issues require testing rather than rhetoric. The eventual results of the modelling studies he proposes should certainly be of interest.

The book is meant to educate everybody. There is a lively chapter of descriptions, mainly of play-fighting, and one giving a brief account of some biological principles. Non-specialists may find it difficult to follow some of the arguments as they dart between assumption, assertion, findings and interpretation. In addition there are overstatements and inaccuracies. For instance, Fagen shows that cost–benefit tables can, in principle, be set up for play with various criteria of "fitness". But there are methodological difficulties in outcome studies that are passed over rather too lightly. The relaxed "play-face", which is a signal for social play in many species, is sometimes observed in solitary object and locomotor play. Fagen treats this as important evidence for the unity of play; but the observations are quite insufficient to decide whether the "play-face" in that case acts potentially even as a sign for others, let alone that it has the same signalling function as in social play. Piaget's theory is outdated; but to state that it considers play merely an epiphenomenon is rather a howler.

In all, there is an air of haste and of conducting a jihad which does a disservice to scholarship and to the subject. Nevertheless, students who already know something about the subject should find the book stimulating. □

Susanna Millar is a Research Officer in the Department of Experimental Psychology at the University of Oxford.

Robert S. Desowitz's *New Guinea Tapeworms and Jewish Grandmothers* (reviewed in *Nature* **296**, 514; 1982) will be published next week in the United Kingdom by W. W. Norton, price £9.25.

nature

17 June 1982

Cloud over French science plan

The French minister for research and technology's plan for future expansion may be delayed by the parliamentary process. But there is in any case a need to think about it more carefully.

The fall of the French franc on Sunday, which brings its devaluation against the German mark to something like 20 per cent in the past 12 months, may concentrate the attention of French parliamentarians wonderfully next week on a key point of the new research law, which is to be debated in the Assembly on Monday: its cost.

The *loi d'orientation et de programmation* for research and development is intended to be the principal instrument of the minister of state for research and technology, M. Jean-Pierre Chevènement. Almost everything hangs upon it: the restructuring of organizations such as the Centre National de la Recherche Scientifique, their stronger linkage with the economy, modifications in contracts of employment and — not least — the outline of a four-year budget. Under the terms of the bill, the government civil research and development budget, effectively Chevènement's budget, would grow in real terms by 17.8 per cent per year on the average as part of a strategy to bring total research and development expenditure in France to 2.5 per cent of Gross National Product by 1985. All basic research is guaranteed at least 13 per cent but high technological priorities (including biotechnology) would get increases of more than the 17.8 per cent. The absolute size of these annual increases is staggering — the best part of one National Science Foundation each year.

This ambitious plan was conceived in the early dreamier days of the present government. Even if the figures are agreed next week at the Assembly, it is doubtful whether they will be faithfully reflected in the 1983 budget, presently subject to horse-trading in the council of ministers. The problem is the overall French budget. The government planned a deficit of FF95,000 million in 1982, but the latest projections give a deficit of FF125,000 million largely because the French economy has not begun to grow at the expected rate of 3 per cent a year, so that tax returns are not covering expenditure. Moreover, increased social welfare benefits and the raising of the minimum wage have pushed up consumer spending but French industry has not been ready to respond, so that consumers have spent their money on imports instead. The government, alerted to the problem, has followed a well-trodden path in France and elsewhere in Western Europe and has imposed a classic squeeze: a four-month freeze on prices and incomes, combined with devaluation. So whither to grand plan now?

President François Mitterrand, speaking only a few days before the squeeze was announced — no doubt to prepare the ground — appeared to be phlegmatic. "We follow the same politics" he said. We are just in its second phase." His government's chief economic objective has been to reconquer the home market. In 1974, 24 per cent of French consumer spending went on imports, but the proportion had risen to 35 per cent by 1981. Part of the government's strategy has been priority for investment in nationalized industry (Mitterrand promised FF25,000 million in 1983) and for industrial innovation. But so far the President has refrained from mentioning a figure for research for 1983.

The research ministry is putting a brave face on this turn of events, but even if the government endorses its hopes for future budgets, trouble in the Assembly about other than financial matters could mean delay. The bill has in any case been badly mauled in the Senate, so that it will be for a joint Senate/Assembly committee eventually to pick up the pieces of Chevènement's grand plan — and to reassemble them as if into a jigsaw puzzle. If the job cannot be done before 1 July, the bill will have to wait until the winter.

The moral in this potentially disappointing tale is general and not specific. The present government of France was elected just a year ago on an optimistic programme of reform and renewal. Neither President Mitterrand nor M. Chevènement has exactly echoed Mr (now Sir) Harold Wilson's British pre-election speech of 1963 about the social benefits of "white-hot technology", but they have both been edging in that direction. The promise is valid, but the performance cannot be guaranteed. This is why the British experience in the 1960s should be studied in France. After the 1964 election, money was indeed thrown at a variety of technical projects but without much lasting effect. (An aluminium smelter built at that time in the Scottish Highlands was shut down only a few weeks ago, the predicted victim of the arithmetic of economics but a social catastrophe nonetheless.) In the end, the inflationary consequences of over-optimism precipitated the stagnation of the British science budget which has persisted ever since (Mr Edward Heath's intervening essay in what is now called Reaganomics notwithstanding). This is why other Europeans are distrustful of what the French have embarked on in the past year. The objective of spending more on research and development is entirely admirable, but nothing will be achieved if the bills have to be paid by printing money. For then it will be recognized that public expenditure on research and development is politically discretionary, and the tap will be turned off.

Who follows Slaughter?

Mr X, the next director of the US National Science Foundation, could spell trouble.

President Ronald Reagan has an unexpected opportunity for putting his own mark on the conduct of US science with the resignation of the director of the National Science Foundation (NSF), John B. Slaughter, whom President Carter appointed in 1980. Slaughter is ending his six-year term of office early, he says, only because he cannot resist an offer to become chancellor of the University of Maryland at College Park, the large state school with a high-quality science and engineering faculty. One of his aims is to integrate the university, which has always been an also-ran to its famous neighbour, Johns Hopkins University, more into the intellectual life of the Washington area. Slaughter says he also wants to attract top quality black students from the region to the university, instead of letting them be lured to other institutions elsewhere in the country. This ambition is understandable. Slaughter is one of the few black scientist-administrators in the United States, and one whose rapid rise has never been tainted with tokenism. (He was an assistant director of NSF, then vice-president and provost of the University of Washington before returning to NSF as director.) But another motive for his decision is the need to pay for his children's college education, and the Maryland job offers not only a pay rise but a free house. Similar reasons were given by another competent Administration figure, Admiral Bobby Inman, when he announced he would leave as deputy director of the Central Intelligence Agency, saying he would have to take a job in business. Must we conclude that the

0028-0836/82/240525-02$01.00

cost of college tuition is now so high that only business executives and university presidents can afford it?

Slaughter denies that being a Democratic appointee in the midst of the Republican Administration has caused problems, or that there is some particular disagreement behind his decision to leave. (The resignation is effective from 31 December.) This denial must be taken with a pinch of salt: the Administration has slashed NSF's science education programmes, particularly close to Slaughter's heart. But it is also true that he worked closely with the incoming Republican team to prevent NSF from being the target of serious cuts in the hard sciences. Is this why NSF has not been the butt of the broadsides of Dr Jay Keyworth, the President's science adviser, against the uselessness of the national laboratories or of planetary science?

In his brief term at NSF, Slaughter has made some important changes, along the lines of his early much publicized call to have NSF sponsor more applied research. He established a directorate of engineering, co-equal with the other directorates that run NSF's basic research. The new directorate, a combination of scattered programmes and some new things, now commands some 10 per cent of NSF's annual budget of approximately $1,000 million. Slaughter has and the National Science Board had a hard job persuading the basic research community to support the engineering directorate. At first, he says, they were "agnostic" about it (read "unenthusiastic"). Basic researchers in the hard sciences tend to regard NSF as their exclusive patron and historically have resisted its forays into public policy, engineering, social sciences and other matters. It is unfortunate, therefore, that Slaughter will not be around to follow through on the move. Past attempts to integrate engineering into the basic research agency have been conspicuously unsuccessful. Unless vigorously and carefully promoted, Slaughter's new directorate could yet founder. Presumably the job of promoting the engineering directorate will fall to Donald N. Langenberg, NSF deputy director, who had the job before Slaughter arrived.

Slaughter has also, with the help of the National Academy of Sciences, established a commission to develop a national policy on science and engineering education, especially urgent in the present crisis in technical education in US elementary and secondary schools (see *Nature* 6 May, p.9). Congress may yet give the commission funds to develop and implement a national plan. Slaughter says that the establishment of the commission is his single most important act as director of NSF. One of his merits is that these changes have been undertaken carefully, in consultation with NSF's traditional constituency, basic researchers in the hard sciences.

Mr X, the highly political candidate whom President Reagan might appoint to succeed Slaughter, could be a very different animal. Reagan is midway through his term of office, and is under fire from his original supporters, the extreme conservatives. If, to placate them, he appoints somebody (probably an industrialist) with fixed and preconceived views about the absolute merits of private enterprise, the result could be disaster. Mr X might well urge that NSF should get out of education policy and social sciences (both hotbeds of liberalism in the extreme conservative view) and no longer give seed money to small businesses doing basic research. NSF could finish up by being stripped down to its core mission — sponsoring basic research in universities — while its other more tentative but important roles, such as its recent brokerage in university–industry partnerships, vanish.

The effect of such a Mr X on the academic science community in the United States is easy to foretell. As a group, academic scientists tend to be paranoid about Washington's designs on their grant monies, and Mr X would strengthen their fears. The community would then pull back from its tentative experiments with industry, its unwilling acceptance, as Slaughter says, that "engineering is also important to the nation", and retreat to its ideological trenches. So it is to be hoped that Mr X will remain just what he is, a figment of the imagination, a hypothetical second-rater, whose *curriculum vitae* may be pressed on the President but whom the President will wisely decide not to appoint to the job Slaughter is vacating.

Nuclear plant and prices

The new chairman of the British electricity industry has a chance to make sense of the nuclear industry.

Dr Walter Marshall, chairman of the United Kingdom Atomic Energy Authority for the past year, is clearly once again the British government's favourite scientist. Last weekend he was made a knight, that peculiarly British public honour which requires that those whom he hardly knows should address him by his first name. A week before that, he was appointed chairman of the Central Electricity Generating Board, the principal owner of commercial nuclear power stations in the United Kingdom and the most likely source of further orders for nuclear plant. Much has plainly changed since 1978, when Marshall as chief scientist at the Department of Energy was for all practical purposes fired by his minister, Mr Tony Benn.

The most obvious change is the change of government (in 1979). But it is also clear that Marshall, whose reputation was previously that of a talented theoretical physicist who had succeeded by deftness and ebullience in carrying the Harwell establishment through hard times in the early 1970s, has used the interval to good effect. His most conspicuous success in the past year has been to persuade the engineers who will be working for

him from 1 July that it is possible to build pressurized-water reactors in Britain at less than twice the cost that obtains elsewhere. The absorbing question now is what Marshall's arrival at the generating board will mean for the British nuclear industry.

The precedents are not uniformly encouraging. Sir Christopher (now Lord) Hinton, who built the first dozen or so nuclear reactors in Britain as well as the separation plant at Windscale (now called Sellafield) maddened his ex-colleagues when, as chairman of the generating board, he insisted that the only significant difference between nuclear electricity and other kinds of electricity is the relative cost. Logically, that view is unshakeable, but it does not thereby follow that the customers (especially monopoly customers) for nuclear plant are without influence on the costs that they must pay. For too much of the past two decades, the generating board has been too passive a customer for nuclear plant, looking to the government to be told when to build reactors, putting up compliantly with the publicly created suppliers' monopoly (called the National Nuclear Corporation) and, more generally, behaving limply.

Marshall, fortunately, is far from limp. The most likely immediate consequence of his translation to the generating board is that this huge nationalized industry will become less dull. The chance of putting the British nuclear industry on a sound commercial footing — or at least of telling conclusively whether such a goal is feasible — is higher now than ever. But in the end what matters is that there should be an asssured supply of electricity at the lowest possible cost. Nuclear power undoubtedly has an important part to play, but so too does the elimination of high cost (oil-fired) and excess capacity from the board's network of generating plant. Pure reason, alas, is not sufficient. Marshall's imaginativeness will help. So too will his capacity to make even those whom he has talked down think well of him.

0028-0836/82/240526-01$01.00

Political interference at FDA?

New policy raises fears of meddling

Washington

For the second time in as many months, the Reagan Administration is being accused of injecting politics into the selection of scientists for independent advisory panels. According to an official at the Food and Drug Administration (FDA), appointment authority over its scientific advisory panels is gradually being taken away from the FDA commissioner and given to the Department of Health and Human Services (HHS). What worries FDA officials is that HHS recently suggested at least two scientists whom FDA believes are unqualified and who were proposed for political reasons.

Earlier this year, the Department of Agriculture admitted that it had been running security checks and "political" checks on scientists nominated for peer-review panels to evaluate proposals for its competitive grants programme but the political checks have since been dropped.

The troubles at FDA, however, seem more serious. Although the secretary of HHS has always had authority to appoint the advisory panels at FDA, that function has traditionally been delegated to FDA itself. One FDA official said that although HHS has given general directions on the numerical balance of the committees; approval by HHS was generally a formality.

Over the past several months, however, HHS has rewritten the panels' charters as they come up for renewal, retracting FDA's delegated control over them. Although FDA is still being asked to identify possible candidates it is also, for the first time, being told to furnish a list of alternatives. And the final decisions are apparently now being made in the office of HHS secretary Richard S. Schweiker.

Of even greater concern to FDA is that the secretary's office has begun to offer suggestions of its own. According to a report in the *Washington Post* last week, which was confirmed by FDA, these suggestions included a woman psychiatrist who was a founder of the "California Pro-Life Council," an anti-abortion group. She was suggested by HHS as the consumer representative on a panel that evaluates contraceptive and abortion drugs. FDA told the secretary's office that she was unqualified, but whether FDA's opinion will be considered is still uncertain.

The *Washington Post* also reported that HHS suggested a California physician whose main qualification, according to FDA scientists, was apparently that he listed Dr Loyal Davis — Nancy Reagan's stepfather — as a reference.

The Department of Health and Human Services claims that it is doing nothing improper. An HHS official said the department is merely continuing the centralization of appointment authority begun under President Carter. The official also denied that the secretary's office was making its own recommendations for committee positions, and said it was simply passing on to FDA all names that it receives from congressmen and senators of both parties.

That interpretation was contradicted by an FDA official, who said that while names had occasionally been passed on before, what was happening now was quite different. The official said that HHS was coming back with new names *after* FDA had submitted its list of candidates, and was specifically asking if these new names were qualified for specific positions. Mr Garret Cuneo, the HHS official now apparently in charge of the selection process, refused to comment on the matter.

The effects of the new appointment policy remain to be seen; no obviously political appointments have been made so far. But even if political manipulation does not occur, FDA is worried that the new procedures will make putting together an advisory panel even harder. The aim in the past has been to get not only leading scientists in a range of specialities, but also to have different geographical regions and women and minorities represented on each panel. FDA officials wonder how these goals can still be met when it is cut out of the process and when it is limited to saying whether the HHS candidates are qualified or not.

Meanwhile, at the Department of Agriculture, the dust seems to have settled. According to the office of Agriculture Secretary John R. Block, the decision to run political checks on scientists nominated for peer-review panels was really the result of a misunderstanding. Political checks are normally run on policy advisers; last autumn, the Office of Management and Budget ordered the department to bring the peer-review panels under the same federal act that governs policy panels. Agriculture department officials then assumed that political checks would have to be run for them as well.

After disclosure of the political checks on scientists, Block ordered them to be stopped. The scientists who had not yet been cleared by Block's office were then all approved at once. Block's office maintains, however, that none of the information

NIH urged to act on germ war

Washington

In response to rumour that the US Army and Navy are seeking to develop "defensive biological weapons systems" using recombinant DNA techniques, two American biomedical scientists have called on the National Institutes of Health (NIH) specifically to prohibit such work.

Dr Richard Goldstein of Harvard Medical School and Dr Richard Novick of the Public Health Research Institute of New York City are proposing an amendment to the NIH recombinant DNA guidelines to ban "the construction of biological weapons by molecular cloning". The proposal will be considered at the next meeting of the Recombinant DNA Advisory Committee on 28 June.

According to Dr Goldstein, several scientists have recently been approached by the military about the possibility of using cloning techniques to produce biological warfare agents. The United States signed the 1982 Biological Weapons Convention, agreeing "never in any circumstances to develop, produce [or] stockpile" biological agents for other than peaceful purposes. But Dr Goldstein says the military apparently interprets this as not covering the development in the laboratory of suitable organisms for "defensive" — by which is meant deterrent — purposes.

In a statement describing their proposed amendment, the scientists argue that "the use of molecular cloning for the deliberate construction of biological weapons is, *per se*, the most serious biohazard imaginable for this technology", and that "it constitutes an egregious misuse of scientific knowledge".

Although the NIH guidelines strictly apply only to researchers working under NIH grants, the Defense Department has so far agreed to follow them. But the amendment's sponsors say that even if their proposed ban is not binding on military research, it would "provide automatic public support for a refusal of the scientific community to participate in the development of biological weapons and it may convince governments that the 1982 prohibition should be construed as applying to laboratory research".

Under the recent relaxation of the DNA guidelines (see *Nature* 29 April, p.793), the previous ban on cloning of toxin genes and on release of recombinant organisms into the environment was lifted, although permission from NIH is required before proceeding with either. The amendment would restore the ban in these two areas when the aim is construction of biological warfare agents, and extend it to previously unregulated activities such as the cloning of viruses for this purpose.

Stephen Budiansky

0028-0836/82/240527-02$01.00

obtained in the political checks was used.

The department will continue to run nominees' names through an FBI security check which it calls "routine", although neither the National Institutes of Health nor the National Science Foundation follows this practice.

Professor Joe Key of the University of Georgia, a director of the competitive grants programme under President Carter, wonders if the checks were an attempt to cause trouble for the four-year-old programme, which is still small and fragile and viewed with some suspicion elsewhere in the department. The bulk of agricultural research (some $1,500 million a year, of which $15 million is for competitive grants) is funded through land-grant colleges and state experimental stations, which receive government funds according to an automatic allocation formula. This latest incident did little to help the competitive programme's vulnerable position.

Stephen Budiansky

West German environmentalists

Greens' delight

Heidelberg

The Hamburg elections on 6 June have left the environmentalist "green" party in a powerful position. While the conservative Christian Democratic Union (CDU) won one more seat in the Hamburg Senate than the Social Democrats (SPD), the greens (Grün-Alternativ-Liste) won nine seats and now hold the balance of power.

In Hamburg, CDU leader Walther Leisler Kiet is now trying to form a non-party "citizens' senate". If he fails, Mayor Klaus von Dohnanyi of SPD will have to ask the greens for their support. The greens have set their faces against coalition and have said that they will cooperate only on strict conditions, including Hamburg's abandonment of nuclear power.

The aftershocks of the Hamburg election continue to reverberate in all political strata in West Germany. The personal intervention of Chancellor Helmut Schmidt and Foreign Minister Hans Dietrich Genscher failed to turn the tide in favour of the Social Democrats, while the Free Democrats, the lesser partner with the Social Democrats in the Bonn coalition, again failed narrowly to reach the 5 per cent threshold necessary to qualify for a seat.

The polarized vote in Hamburg reflects the shakiness of the Bonn coalition as well as the strength of the green-left. If the Social Democrats lose only one more *Land*, the Christian Democrats will have a clear two-thirds majority in the Bundesrat and could block all government legislation. Rumours are now rife that the Free Democrats will leave the coalition after the Hesse election on 26 September, perhaps joining with the Christian Democrats. The deepening green intrusion into West German politics is at least partly responsible.

With seats in five parliaments, the greens have become the third strongest party. The abrasive challenge they offer to the assumptions of the established social industrial system is proving hard for even the Social Democrats to meet. They are concerned that "integrating" green issues would further alienate the traditional working class base without necessarily keeping ecologically-minded young voters. Indeed, there are wide differences between the Social Democrats and the greens except on individual issues such as the Rhine-Danube canal or the protection of wildlife.

The central goals of the greens are the realization, by means of a social welfare state, of the recommendations of Global 2000 and of disarmament. They call, for example, for radical change to a no-growth society, the use of alternative technologies and energy sources, conservation, economy and peace. They oppose the "compromise politics" of the Social Democrats.

Voters are mainly young and well educated and the focus is often local: cauliflower cancers of eels from the polluted Elbe and the discovery of tonnes of nerve gas in a ramshackle chemical factory near the city centre were important at Hamburg. The parties are autonomous at every level, so that there are no declared leaders and the small new national party controls neither the policies nor the attitudes towards coalition of the *Länder* parties. To ensure the five per cent vote, the Länder-Grünen have often joined with other groups, some of whom have made uncomfortable allies.

Their anarchic demands for Rousseauesque self-determination, their use of violence and frequent threats to make West Germany ungovernable if minority demands are not met raise spectres of the Weimar Republic.

In September, the Hessen greens will go it alone, campaigning especially against the extension of Frankfurt Airport and plans for four local reprocessing plants for atomic waste. They expect a 10 per cent vote, a Christian Democrat parliament and a defeat for the Social/Free Democrat alliance that would indeed have national repercussions.

Sarah Tooze

European science policy

Give and take

Brussels

The second report on European science policy by Crest, the Committee on Scientific and Technical Research, tells the familiar story of how, after 15 years of trying to bridge the technology gap with the United States, Europe still lags behind in the latest technologies. Of the million or more European scientists — representing 20 per cent of the world's total — 350,000 are researchers, and science in Europe cost $35,000 million in 1980. The new report, which has still to be published, goes beyond its predecessor in comparing the national research and development policies of the ten EEC member states.

Energy research, which since 1974 has been a priority, accounted for a smaller share of the research and development budget at 11 per cent in 1980 than in either the United States (12 per cent) or Japan (25 per cent). Nuclear (fission) research has seen its importance slump, particularly in West Germany and France.

Allocations to industry research grew by 15 per cent a year in the Community as a whole between 1978 and 1980 but are thought to have returned to pre-1974 levels only last year. The increase has been marked in Italy and the United Kingdom, but UK spending is still below the 1974 level.

Space research, where international cooperation is at its most developed in Europe, shows France and West Germany sharing more than two-thirds of the spending, worth $860 million in 1980. Curiously Italy and Belgium, recent years have contributed the greatest proportion of their budgets to space research.

Agricultural research turns out to play the biggest role in the smaller member states. In Ireland and Greece a quarter of public research and development money goes on agriculture, which is equivalent to only 3.7 per cent of the EEC's total spending. Agriculture is also important in Denmark and the Netherlands which spent $3.7 and $6.8 per head respectively compared with an EEC average of $2.7, and $3.1 in the United Kingdom. Italy's outlay seems small at $0.9 per inhabitant, considering agriculture's social and economic importance in the country.

International cooperation is still important but its share of public budgets has fallen by as much as 25 per cent since 1978. Between 1974 and 1978 it was constant at 10 per cent but now there are wide differences among the member states. Britain's contribution, for example, has increased by 25 per cent since 1970.

The ten members of the EEC also differ in the response of their companies to the technological challenge. The role of the state in industrial research is greatest in the United Kingdom (30 per cent), France (22 per cent) and West Germany (20 per cent), but less than 10 per cent in the rest. Measuring research in companies as the proportion of expenditure on research and development to the value added of the manufacturing puts British companies first. The proportion of research expenditure is 4.7 per cent followed by West German companies (3.9 per cent), French companies (3.5 per cent) and Dutch companies (3.3 per cent); in the United States though, the average is 5.7 per cent. But if a distinction between public expenditure and companies' own funds is made, the picture changes. US companies spend the most (3.8 per cent) followed by the Japanese (3.4 per cent).

The report argues that many of the

0028-0836/82/240528-02$01.00

member states are not pulling their weight. Germany, France and the United Kingdom share 80 per cent of Community research and development expenditure while the combined efforts of Belgium, Denmark, Ireland and Greece amount to as little as 6 per cent. The conclusions are being studied in preparation for the next EEC science council of ministers on 30 June.

Jasper Becker

Monsanto hands out $23.5 million

St Louis

The Monsanto Company has awarded $23.5 million to the medical school at Washington University in St Louis, Missouri, for five years of research into how proteins and peptides affect cell regulation. The grant, announced on 3 June, approximately doubles the research funds from non-federal sources at the medical school for that period, and is one of the largest from a single company to a university.

Monsanto's award is an obvious boon to the university's research programme: two-thirds of the money will go to applied research and one-third to basic research. Work will be selected by an eight-member committee — four university faculty members and four people from Monsanto — headed by Dr David M. Kipnis, head of the department of internal medicine at the university medical school. University researchers will be at liberty to publish the results of work funded under the grant, but if the material contains potentially patentable technical developments Monsanto can review it, and request a short delay before submission for publication. Monsanto will have exclusive rights to any licences arising from patents on the work, but patents will be the exclusive property of the university, which will be able to receive royalties from Monsanto licences. Royalties will go to the university's research and education programmes — not to individual researchers.

Meanwhile, as many as "a couple of dozen" Monsanto scientists may work at the university and some may spend several years there. Monsanto's move clearly shows an interest in medical products, and Dr Howard A. Schneiderman, Monsanto's senior vice-president for research and development, says that while the company does not now market health-care products, it hopes to do so in future.

Monsanto launched a $1.8 million project with Washington University earlier this year for research using monoclonal antibodies for diagnostic work. The company is also working with Dr Mary-Dell Chilton, of the university biology department, on genetic engineering in plants. Schneiderman did not, however, confirm rumours that a major new grant from Monsanto was imminent to support Chilton's research.

Karen Freeman

Luxembourg in space

TV plan delay

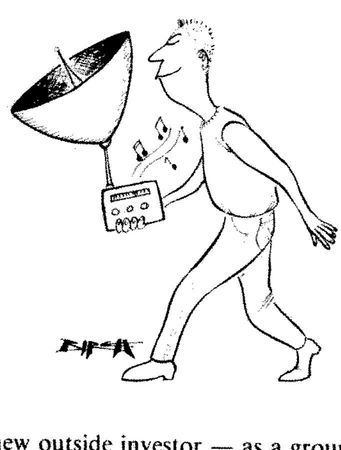
Luxembourg is falling behind in the race to provide the first satellite broadcast for Europe. Unresolved wrangles about the relative sizes of French, Belgian and West German shareholdings in the projected £200 million Luxembourg system have now put back the earliest possible launch date until late 1986.

By then, the national satellites planned jointly by France and West Germany will probably have been up for a year. Also by late 1986, the new all-British Unisat, with two BBC channels, might be aloft.

Radio Tele-Luxembourg (RTL), the commercial company in charge of all broadcasting in Luxembourg, insists that its satellite project is a matter of when, not if. It plans to beam three channels of films, news, entertainment and advertising to a wide swathe from Benelux to Bavaria, and possibly to south-east England. It knows, however, that this ambition represents a big gamble for a small country. RTL, whose current radio and television programmes reach a weekly audience of 40 million in Europe, contributes 5 per cent (about £25 million) of the annual income of the Luxembourg government. Commercial broadcasting has been a major source of revenue for the 50 years since Luxembourg, accustomed to living on its wits, recognized that its central location, plus national sovereignty over its broadcasting policy, permitted it to send radio advertisements and popular music into countries whose governments prohibited one or both by their own broadcasting organizations. It was in 1932 that Radio Luxembourg began spoiling the determination of Lord Reith, the BBC's first director-general, that the British people should have no light radio entertainment on Sunday.

Now, however, with commercial radio and television sprouting all over Europe and eroding its audience, RTL must find a new service to sell or see its profits and contribution to the national purse atrophy. But even if its satellite succeeds in pulling in audiences and advertising revenue, the heavy costs of depreciation of the start-up years is going to hurt Luxembourg's economy, already in trouble with its first unemployment since the Second World War.

RTL cannot proceed, however, until it decides upon a new financial structure to accommodate the necessary infusion of new capital. Although the company is registered in Luxembourg and has, by law, a majority of Luxembourg nationals on its board, its largest shareholder is predominantly Belgian. Also, French investors, notably in the form of Havas-IP, Compteurs Schlumberger and Paribas, are in a strong if not controlling position. RTL's problem now is how to introduce a



big new outside investor — as a group of West German publishers (which does not include the giants) has offered about £50 million — without disrupting the delicate Franco-Belgian balance of power. The French hope to avoid excessive dilution of their own influence. But they, and the whole board, know that the matter will have to be sorted out next month if further delay is not to occur.

Once the go-ahead is given, RTL will invite tenders for the system (two satellites plus a spare). It will also decide whether to devote one of the three television channels to broadcasting in English. Any deregulatory move by the Hunt committee now inquiring into the expansion of cable television in Britain would undoubtedly sway the decision. Elsewhere, RTL insists, a vast expansion of cable television in Europe is not necessary to the success of its plan. It expects half of its audience to buy a rooftop dish to receive its programmes.

Brenda Maddox

Jobs and automation

US faces facts

Washington

The effect that new electronic technology is having on jobs, which has been an issue of intense debate in Europe, is now a cause for concern in the United States.

A new study by the government's General Accounting Office (GAO) suggests that the recent revolution in electronics will be felt not just in manufacturing — the sector most affected so far by automation — but also in office and service jobs. And even within manufacturing, the spectrum of fixed automation introduced in the 1950s, and the initial applications of robots lately have tended to be in tasks that were considered menial, or monotonous or unsafe, for human workers. Spray painting

0028-0836/82/240529-02$01.00

of automobile bodies is an example. Now, with microprocessors that can be incorporated into virtually every machine, even skilled and previously immune occupations such as tool-and-die making may be affected.

The critical question is the validity of what has been a traditional assumption: that technological progress brings with it more jobs. Several witnesses scheduled to testify next week before a House subcommittee investigating the issue doubt that assumption still holds. "We can't count on expansions in the white-collar or service areas, which is what saved us in the fifties and sixties", says William Bittle of the International Association of Machinists and Aerospace Workers.

An annual employment forecast issued by the Bureau of Labor Statistics (BLS)* confirms that employment in at least some office and service occupations is being hit by electronic technology. "We do see some jobs disappearing", says Ronald Kutscher, assistant commissioner of BLS for economic growth and employment projections. Key-punch operators, telephone operators and virtually all the printing trades will be hard-hit, for example.

The other open question, though, is how many jobs will be created by the new technologies directly. Workers will be needed to build, install, adjust, and repair automated equipment. The GAO study found virtually no evidence that could answer this question, however.

The trade unions have apparently accepted that jobs will be displaced by automation. But the critical issue to them is whether enough time will be allowed for workers to find new positions. In Norway, unions have negotiated contracts that set a gradual rate for the introduction of new technologies. The possibilities of such contracts being agreed to in the United States seem much smaller. A common complaint by American trade unions is the tendency towards secrecy on the part of management and the absence of the sort of cooperation and consultation practised in Europe and Japan.

The House subcommittee hearings may be a small step towards some government action on the problem. Representative George Miller, who is holding the hearings, has introduced a bill (HR 5820) that would provide for vocational retraining of displaced workers in new occupations created by automation. The unions, however, tend to dismiss government-supported training as a subsidy for industry and an inefficient substitute for on-the-job training. More to the point may be another concern of Miller's: he points out that the government spends nearly \$2,000 million a year on labour-saving devices. **Stephen Budiansky**

Advances in Automation Prompt Concern Over Increased U.S. Unemployment (General Accounting Office, May 25, 1982). *Occupational Outlook Handbook* (U.S. Department of Labor Bureau of Labor Statistics, April 1982).

GM cancer prizes

Rules to be bent

Although this year's General Motors Cancer Research Prizes have been duly awarded (see below), leaving Dr Howard Skipper, Dr Denis Burkitt and Dr Stanley Cohen each \$100,000 better off, the awards committees are clearly running into difficulties in selecting an annual trio of winners while sticking to the rules. Only four years after the awards started, the biggest worry is that of finding each year someone worthy of the prize "for the most outstanding recent contribution to the prevention of cancer, including environmental factors".

The rules of the prizes were set in 1978 when General Motors, disturbed by the number of its directors who had become victims of cancer, put \$2 million (just doubled) into a General Motors Cancer Research Foundation. The prizes are large enough to invite comparison with the Nobel awards; the rules, however, differ in interesting ways.

One rule, intended to eliminate fortuitousness, is that a prize winner should have made more than one major discovery. Their discoveries must have been made within the previous fifteen years unless their importance has been recognized only more recently.

One prize (Kettering) is for diagnosis and treatment of cancer, another (Mott) for prevention and the third (Sloan) for a contribution to basic science. Winners are chosen by a process that resembles that used for the Nobel prizes. From a list of 25,000 prominent scientists, about 6,000 each year are asked to nominate candidates. Three subcommittees, one for each prize, first pare the nominations to twelve. Last year, they had to sift through 114, 40 and 91 nominations respectively. By the second meeting, each committee member has to report on the merits of two of the twelve candidates, eight of whom are then eliminated. At a final meeting the committees rank two of the four remaining candidates in order of preference. Finally the awards assembly has to decide whether to follow the committee's advice.

This year the assembly argued whether it should award the Mott prize for prevention. Nobody seems to have doubted the importance of Dr Burkitt's discovery of the childhood cancer that now bears his name (Burkitt's lymphoma) and his perceptive suggestion that it is transmissible (it later became clear that a virus is involved). Nor is it in doubt that he pioneered the chemotherapy of "his" lymphoma. But that was all more than fifteen years ago and in any case cannot strictly be considered a contribution to the prevention of cancer.

Turning a blind eye to those problems, the relevant committee and the assembly also had to grapple with the question whether Dr Burkitt's advocacy of the importance of dietary fibre in the prevention of cancer, the topic that has most occupied him in the past fifteen years, is more than a provocative hypothesis. In the end, it was not taken into account.

The choice of Dr Howard Skipper for the Kettering prize for diagnosis and treatment ran into much less opposition, although again the rules have obviously been stretched. Skipper is widely acknowledged as a pioneer of cancer chemotherapy. For 35 years he has influenced clinical chemotherapy by extensive studies on animal and cell models. His discoveries have influenced which drugs are used, in what combinations and their dosage and timing. It is, however, not easy to point to two major discoveries of Skipper's within the past fifteen years. His most recent work bears on the understanding of drug resistance in tumours.

Even for the least disputed of this year's prizes — that to Dr Stanley Cohen — an elastic interpretation of the rules is evident. There is no doubt that he put epidermal growth factor on the map and that it is relevant to cancer research. Cohen's earlier and very important work on nerve growth factor is not a contribution to cancer, and so the characterization and the biological effects of epidermal growth factor have had to be considered separate discoveries.

Perhaps prize rules are made to be stretched. Certainly as Robert Burton once said: "No rule is so general, which admits not some exception". But when the exception is the rule it may be time to change them. **Peter Newmark**

General Motors Cancer Research Prizes 1982: From left to right: Professor Stanley Cohen of the department of biochemistry at Vanderbilt University School of Medicine; Dr Howard E. Skipper, recently retired president of Southern Research Institute, Birmingham, Alabama; and Dr Denis Burkitt, honorary senior research fellow, St Thomas's Hospital, London.

0028-0836/82/240530-01\$01.00

Awaiting the gypsy moth

Washington

Spring in New England, with its green-shaded campuses, tall shadowy and seemingly endless deciduous forests, and country villages with white church spires set against rolling green mountains, used to be one of the natural glories of North America. But as spring advances in New England this year, so will the gypsy moth (*Lymantria dispar*). During June and July the moths will emerge from their pupae and begin their destructive predatory march that now covers most of eleven states from Maine down to Maryland, munching away at the entire foliage of white, red, black and scarlet oaks, and in infested areas of the grey birch trees, the apple trees, American beeches, red maples and even conifers including the common white pine. In areas of heavy infestation, the upper ridges of mountains become so entirely denuded of tree canopy and ground cover that they are ashen and apparently lifeless, reminiscent, as one New Haven scientist said last week, of an experiment that Brookhaven National Laboratory ran in the 1950s to see what a forest would look like after irradiation by nuclear weapons.

The plague of the gypsy moth in the United States is a genetic experiment run wild. In 1869, L. Trouvelot, a French astronomer and naturalist who lived at 27 Myrtle Street, Medford, Massachusetts, imported some gypsy moths from Europe to help found a silkworm industry in New England, then the centre of the textile industry. Some of the eggs were lost, and by 1889 they had multiplied enough to provoke an outbreak in the town of Medford. The following year the state legislature awarded $25,000 for control.

Things have got worse since. The larvae hang on silk threads they have spun, so that air currents can carry them to new trees in which to hatch. Prevailing winds encouraged the spread through northern New England. In 1923 the government established the north–south flowing Hudson River as a natural "barrier zone" to contain the spread south. The Second World War interrupted control efforts, but brought an apparent remedy: DDT. The 1950s saw uncontrolled spraying which checked the infestation in some places, but did not control the spread. In 1958, when DDT was banned, a less effective substitute Sevin was introduced and the government finally started a research programme to control the beasts.

Each year, the front of infestation moves westwards and southwards away from New England. Last year an estimated 13 million acres were defoliated, double the area in 1980. Estimates for the summer of 1982 are that yet another 13 million acres will be defoliated, that is, 50 per cent or more of the canopy will be destroyed by gypsy moths. Carried by vehicles, gypsy moths have been found in Ohio, North Carolina, South Carolina, Virginia, West Virginia, Wisconsin and Michigan and outbreaks have been reported in Salem, Oregon, and Santa Barbara, California. New England forests are not lumbered extensively any more, but as forests in the south and west become infested, the economic costs of the gypsy moth plague will grow.

The battle against the gypsy moth has been hindered by two things. One is the fact that no major industry is affected — thus in Maine, for example, the localized infestations are less important politically than the spruce budworm infestation, which has hurt that state's timber industry for some years. A second impediment is that the US environmental movement, with its power to express issues forcefully, and arouse public opinion, has been lukewarm to the gypsy moth issue. Many environmentalists are more concerned about the potential toxicity of Sevin, a spray commonly used against the gypsy moths, than they are about the forest damage *per se*. Moreover, the fact that lesser infestations do not destroy a forest completely and that most of the trees affected are only "ornamental", gives the environmentalists grounds to argue that the problem is not so urgent. One suspects that they may also hesitate to decry the gypsy moth because if the plague were indeed extensive is it not only a few logical steps to argue that we should return to the only known substance that combats it — DDT? Many environmentalists may not want to start down that slippery road.

So far, the gypsy moth story is one of the failure of science and technology. As the old saying goes, if we can go to the Moon, why can't we beat the gypsy moth? The chief microbial pathogen used so far, BT (*Bacillus thuringiensis*), is not nearly as effective as DDT: it must be sprayed over a large area requiring that an entire neighbourhood should get together and spray at once; if rain or other weather changes interrupt the spraying, it does no good. The other remedies tried so far have been of little effect, partly because, by now, the infestation is so severe. Research has recently focused on pheromones and other specific chemical lures, but the results are not yet broadly applicable.

The biology of the gypsy moth raises several interesting biological questions. The population dynamics of the gypsy moth are not understood, with rises of population followed by collapse, perhaps because of the weather. The physiology of the single American species may throw light on its susceptibility to parasites, perhaps the best hope for control. The relative scarcity of *Lymantria* species in Europe as compared with East Asia suggests that the genus originated in the Far East. The hope is that the work now under way will suggest a species-specific pesticide for a pest that the US Department of Agriculture acknowledges to be uncontrollable. The chemical company that first finds a cure will make a fortune. The search for parasites effective against gypsy moths has now turned to East Asia.

Upland oaks in New England heavily defoliated by the gypsy moth.

The pest also haunts Europe, but because the forests are less dominated by oak and are smaller and more precisely managed, it has not become the plague it is in the United States. Moreover, in Europe, the parasites that limit the gypsy moth population survive the winter whereas, when introduced into the United States, they die for lack of a suitable winter host.

China is the only part of the world that has not been thoroughly investigated for parasites that could solve the problem in the United States. It is to be hoped, therefore, that the Chinese government will be as cooperative as possible with the teams from the US Department of Agriculture which are there now, and plan to go again next year, to investigate gypsy moths in China, particularly in allowing parasites to be taken out of the country.

It is also to be hoped that the Reagan Administration will produce a better research policy on the gypsy moth than its parsimony so far suggests. The research programme begun by the Department of Agriculture in the 1970s may for practical purposes be abandoned as a consequence of the budget cuts. Aid to affected states on research and control, to which the federal government used to contribute 43 per cent, was cut for the present fiscal year to 12.5 per cent, although the Administration relented three weeks ago, and raised the proportion to 25 per cent. But in the coming financial year the Department of Agriculture will not be allowed to embark on any joint programmes with the states, although it may continue with some research of its own.

As New Englanders sit on their porches this spring, musing on how much the gypsy moth will devour of scrub oak or aspen trees on the West Coast, they can take some morbid pleasure in reminding themselves that their President thinks this is a local problem, and that his science adviser, for all his talk about the importance of science's interface with society, does not think the gypsy moth is of any priority for science.

Deborah Shapley

0028-0836/82/240531-01$01.00

CORRESPONDENCE

Herczynski's arrest

SIR — Many of your readers will have been dismayed to learn from Vera Rich's note (*Nature* 20 May, p.172) that Dr Ryszard Herczynski has been arrested in Warsaw for passing to two attachés from the US Embassy "materials damaging to the interests of the Polish state". Herczynski's numerous friends in this country know him as a vigorous and fearless critic of attempts by the authorities in Poland to inject political considerations into the conduct of scientific and academic affairs. That he has been critical of action taken under the present martial law, we can be sure; and it is probable that the materials alleged to be damaging to the interests of the Polish state were simply documents expressing such criticisms. Nevertheless, he could be sentenced to between three and five years' imprisonment (not internment) under martial law, with no appeal. This would be a terrible injustice.

Expressions of support for Herczynski from Western scientists may dissuade the Polish government from treating him harshly. It would be helpful to convey to those concerned that the essentially non-political causes defended by Herczynski — respect for the truth in scientific and academic matters and maintenance of the integrity of scientists — are also our causes, and that we do indeed care what happens to him. Quick action is desirable since Herczynski's trial may be held soon.

A letter of protest has been sent to the Polish authorities, signed by 25 scientists.

G.K. BATCHELOR

Department of Applied Mathematics and Theoretical Physics,
Cambridge, UK

Japanese IQ

SIR — Reading Alun Anderson's comments on the study of IQ in Japan (*Nature* 20 May, p.180), the surprising find is not that it has improved with better nutrition, urbanization, improved social conditions and so on, but that under conditions comparable with those in the United States and Europe, it is 10–15 per cent higher. If this difference is genuine, I could offer an easily testable explanation.

It is well known that coaching can improve IQ results by 5–10 per cent. My conjecture is that every Japanese child undergoes a sort of coaching when (s)he learns to read and write his own language. Apparently it is so complex that only by the age of eleven can they be regarded as fully literate. Of course, the ideal IQ tests are language neutral, but my point is that the mastering of a difficult language plays a general IQ boosting role. This hypothesis is easily testable: one should compare the IQs of immigrant Japanese groups in the United States, say, who are literate in their language and matching groups who are not. Also, if the hypothesis is true, other members of the Chinese-related cultures should show higher IQ, under similar social and economic conditions.

Finally, competent linguists could perhaps find other languages of comparative complexity, the speakers of which could then be included in such inter-nation IQ comparison.

G. MAGYAR

Abingdon, Oxon, UK

Drugs and safety

SIR — I assume that the attack by M. Weatherall (*Nature* 1 April, p.387) on the toxicological testing of drugs was designed to provoke a response and was not intended to celebrate its publication date.

To propose the substitution of patient surveillance for toxicological testing is at first sight attractive because it places the emphasis firmly on men and women and not rats or bacteria. There is, moreover, undoubtedly a kernel of truth in some of his concern about unthinking and invalid testing, yet in selecting genetic toxicology to bear the brunt of his criticism he has chosen unwisely. His arguments show ignorance of current attitudes among thinking toxicologists and regulators. Moreover, the effects under consideration are, as I shall show, almost completely refractory to conventional surveillance.

Genetic toxicologists aim to detect potential mutagens and initiators of carcinogenesis. They would not expect that the result of a bacterial mutagenicity test alone could be extrapolated directly to man. Properly thought-out tests carried out early in the development of a drug can give an alert to possible problems ahead, problems that may be averted by chemical modification or that may require specialized tests to be undertaken when the drug is first given to humans. It would, of course, be verging on the criminal to give young patients with benign disease a drug known to be a potent mutagen in a wide variety of systems (including whole animal tests). Nevertheless, weak mutagens, or those active only in certain rather special situations, may well have a clinical future when all relevant circumstances have been considered. It is here (at the regulatory level) that the common sense advocated by Weatherall should be seen to operate.

To argue for surveillance and against testing for mutagenicity is particularly inappropriate since the deleterious effects of DNA damage are in general so long delayed that they would fail to be detected. Although sensitive techniques are now being developed that would enable the presence of DNA damage and other markers indicative of DNA damage to be detected in treated patients, their use is justifiable on economic grounds only where mutagenicity tests indicate possible risk.

Weatherall is right to try to educate the public and their regulatory bodies not to expect absolute safety from drugs any more than from surgical operations, but the surgeon introducing a new technique will have spent many hours experimenting on animals and will have a fair idea of possible risks before his first operation on a patient.

B.A. BRIDGES

MRC Cell Mutation Unit,
University of Sussex, Brighton, UK

Back to basics

SIR — Many American scientists may be unconvinced that "Europe leads on (nucleotide) sequences" as suggested by your correspondent Robert Walgate in his article on the European Molecular Biology Laboratory (EMBL) computer library of nucleic acid sequences (*Nature* 15 April, p.596). American molecular biologists have for several years been successfully using several similar data banks including the Nucleic Acid Sequence Data Base organized by Margaret Dayhoff's group at Georgetown University Medical Center. This sequence library, in addition to containing several programs for sequence analysis, currently includes 746,000 nucleotides (compared with the 600,000 reported in your article for the EMBL library).

Apart from the issue of nucleotide quantity, the sample entry provided in your article, while perhaps not typical, does not reflect favourably on the EMBL library as regards quality. The entry gives no indication that the MOPC41 kappa gene sequence presented was not determined by the authors of the entry reference; the sequence was in fact copied by them in their paper to compare with a sequence of another gene which they had determined. In transcribing the MOPC41 sequence these authors erroneously inserted a nucleotide (position 189) which now appears in the EMBL sequence. This trivial error illustrates the importance of relying on original sources of sequence data, a policy of the Georgetown data base. The correct sequence is included in the corresponding entry from the Georgetown data base (see below) along with the correct primary reference (Seidman *et al. Nature* **280**, 370; 1979).

JONATHAN SEIDMAN
EDWARD E. MAX

Department of Health & Human Services,
National Insitutes of Health,
Bethesda, Maryland, USA

```
KVMS41
Ig kappa chain V region germline gene VK41 - Mouse

Seidman, J.G., Max, E.E., and Leder, P., Nature 280, 370-375,
    1979 (Residues 1-664)
     Residues              Feature
     120-174,              Protein: Ig kappa chain precursor V
     303-598               region MOPC 41
     175-302               Intron

Composition:   168 A,   153 C,   139 G,   204 T
Length:        664

             10         20         30         40         50         60
  1 CGTGACCAAT CCTAACTGCT TCTTAATAAT TTGCATACCC TCACTGCATC GCCTTGGGGA
 61 CTTCTTTATA TAACAGTCAA ACATATCCTG TGCCATTGTC ATTGCAGTCA GGACTCAGCA
121 TGGACATGAG GGCTCCTGCA CAGATTTTTG GCTTCTTGTT GCTCTTGTTT CAAGGTTAAA
181 ATGAAACTAA AATTGGGAAT TTTCCACTGT TTCCAACTGT GGTTAGTGTT GACTGGCATT
241 TGGGGGATGT CCTCTTTTAT CATGCTTATC TATGTGGATA TTCATTATGT CTCCACTCCT

            310        320        330        340        350        360
301 AGGTACCAGA TGTGACATCC AGATGACCCA GTCTCCATCC TCCTTATCTG CCTCTCTGGG
361 AGAAAGAGTC AGTCTCACTT GTCGGGCAAG TCAGGACATT GGTAGTAGCT TAAACTGGCT
421 TCAGCAGGAA CCAGATGGAA CTATTAAACG CCTGATCTAC GCCACATCCA GTTTAGATTC
481 TGGTGTGCCC AAAAGGTTCA GTGGCAGTAG GTCTGGGTCA GATTATTCTC TCACCATCAG
541 CAGCCTTGAG TCTGAAGATT TTGTAGACTA TTACTGTCTA CAATATGCTA GTTCTCCTCC

            610        620        630        640        650        660
601 CACAGTGATA CAAATCATAA CATAAACCCC ATGAAAGTAG AAATGAGAGG CTGGGCTGCT
661 CTGA
```

NEWS AND VIEWS

Female mate choice and runaway sexual selection

from Paul H. Harvey and Stevan J. Arnold

IN *The Descent of Man and Selection in Relation to Sex*[1], Darwin explained how various traits usually found in only one sex — such as the peacock's tail or the deer's antlers — might have evolved. He thought that such traits resulted from one of two selective processes that favour access to mates: either selection for weapons used in combat between males, or selection for adornments used to attract females (Figs 1, 2). Darwin observed that the adornments are most clearly seen in males of species, for example, birds of paradise, where females choose from a group of displaying males. But Darwin was aware of a difficulty in his theory that has never been fully resolved: the characters favoured by females often appear to hinder the survival of males. Why should females choose to mate with those males least likely to be favoured by natural selection? Recent analytical work formalizes a model suggested by R.A. Fisher and shows how selection can favour female mating preferences for characters that actually lower male viability, even to the point of driving a population to extinction.

Fisher[2] described a simple model for female choice which had a surprising outcome — male characters which impair survival could be selected for and maintained in a population simply as a consequence of female choice. He considered species with heritable variation for both a male character, say tail size, and female preference for choosing mates on the basis of tail size. If environmental change led natural selection to favour an increase (or decrease) in male tail size, then, over the generations, male tail size would change. However, so would female preferences. The important idea here is that males with longer tails would tend to be mated by females who prefer longer-tailed males. The male offspring of those females will not only inherit their fathers' long tails and be favoured by natural selection but will also inherit their mothers' preference genes and pass them on to their daughters, thus causing an increase in the frequency of those preference genes. This genetic covariance or linkage disequilibrium is a result (not an assumption) of Fisher's model. Modification of male tail size will, therefore, proceed under the influences of both natural selection and sexual selection.

Ultimately a balance between the forces of natural and sexual selection might be expected to lead to an equilibrium male tail size, but Fisher[2] argued that "as long as there is a net advantage in favour of further plumage development, there will also be a net advantage in favour of giving to it a more decided preference". Fisher termed this process 'runaway selection' and went on to assert that in the absence of "severe counterselection . . . *it is easy to see* that the speed of development will . . . . increase with time exponentially, or in geometric progression" (our italics). He suggested that the process might only be checked when males had tails so long that they would be extremely unlikely to survive and females preferring to mate with them would be left without a mate.

Fisher never published a formal population genetic model of the runaway process and, since his writing style was never lucid, later workers have often chosen to add assumptions to the model. One common variant[3,4] claims that the runaway process is a consequence of females choosing males on relative rather than absolute criteria — for example, from one extreme of the distribution of observed male tail sizes in the contemporary population rather than males with particular sized tails.

There has been a clear need for explicitly formulated models that would describe the joint evolution of male characters and female choice and that would, at the very least, specify some conditions under which the runaway process does or does not operate. In a series of papers published in the 1960s and 1970s and summarized in a book[5] published in 1980, O'Donald examined several models which incorporated genetic loci coding for male characters and female choice. These two-locus diploid systems cannot be treated analytically, but computer simulations show that the runaway process would occur. However, Fisher "describes the

Fig. 1 (right) Strepsiceros Kudu. One category of sexually selected characters confers a direct advantage in male–male combat over access to mates. Darwin was aware of numerous examples and had no difficulty in explaining them. Male weaponry is often more markedly developed among polygynous species. **Fig. 2** (below) A second category of sexually selected characters is particularly evident in males of lek species, such as the bird of paradise, where females choose from among a group of displaying males. The males play no part in defending the nest or feeding the young. But why do females appear to choose to mate with males whose plumage must be disadvantageous in terms of natural selection?

Paul H. Harvey is a lecturer in Biology at the University of Sussex Falmer, Brighton BN1 9RH and Stevan J. Arnold is an Associate Professor of Biology at the University of Chicago.

0028-0836/82/240533-02$01.00

evolution of mating systems in terms Darwin himself might have used"[5] and not all his claims are validated in O'Donald's population genetic treatment of his model — the runaway process does not increase at a geometric rate.

Working with a haploid two-locus analytical model and diploid two-locus simulations, the recent treatments by Kirkpatrick[6] also demonstrate the runaway process. But it is evident from Kirkpatrick's treatment that both Fisher and O'Donald were wrong to conclude that "preferences start to evolve only if females prefer those male phenotypes that are advantageous in natural selection"[5]. Indeed, female preference genes that favour more viable males do not necessarily spread through the population, and neither do the more viable males! Why should this be so? The key to the answer lies with Kirkpatrick's finding that when the male trait gene frequency is plotted against female choice gene frequency then there is a line of genetic equilibria: for every frequency of preference there is a frequency of the male trait at which the system is in equilibrium. There is not a single equilibrium point because any viability loss experienced by long-tailed males can, with some female preference, be exactly balanced by a mating advantage. When females preferring long-tailed males are common in the population then sexual selection will maintain a high frequency of the long-tailed males. When the preference gene is not at fixation, however, a low frequency of shorter-tailed males may be maintained in the population to service those females who prefer mating with them. The process is clearly frequency dependent.

What this means is that when the same male fitness and female choice functions are used while initial trait and preference frequences are varied then different points on the line of equilibria are reached. Once populations reach the line there is a prospect for additional indeterminate evolution — genetic drift can carry small populations along the line of equilibria. Finally, it is clear from the models that females need not choose males on relative rather than absolute criteria (see above and ref. 7) for the male's tail to evolve to a maladaptive size.

Kirkpatrick's analysis independently followed a treatment of the problem by Lande[7] which, more realistically, assumed that both male traits and female mate preferences are under polygenic control. Lande modelled such a system with male traits and female preferences treated as normally distributed variables. Although we might expect selection to deplete heritable variance in both traits, Lande points out that this is not necessarily so[8]; at some point depletion by selection will be balanced by enrichment of variation by mutation and recombination. If we assume that variances in male trait and female preference metrics remain the same

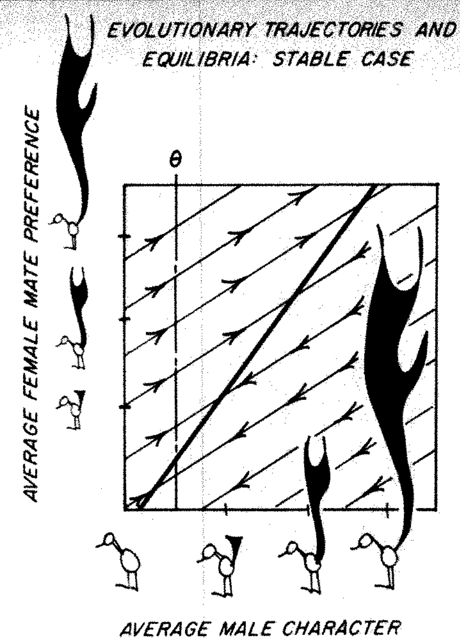
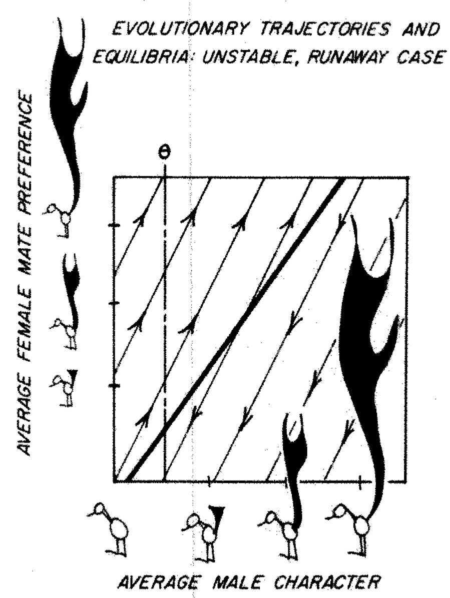


Fig. 3 Females choose mates on the basis of tail size. The light lines depict possible trajectories for the joint evolution of male tail size and female preferences, and the arrows indicate the direction taken. (The male tail size that is optimal for viability is indicated by the symbol θ.) The heavier line connects the equilibrium states at which the males tail length and the females mating preference cease to undergo evolutionary change (other than by random drift along the line). Whether the evolutionary process leads to stable or unstable equilibria depends on genetic parameters and the strength of viability selection.



between generations, but that character averages change as consequences of natural and sexual selection, then Lande's models also reveal a line of equilibria. For any strength of female preference there is a male trait size such that the system is in equilibrium. Lande demonstrates the line of equilibria in three models assuming quite different rules of female choice. Again, there is no direct selection on female mating preferences, which evolve as a genetically correlated response to selection on males.

Lande's formulation allows a different interpretation of what Fisher meant by runaway selection. With the two-locus models, populations perturbed slightly from the line of equilibria always return there (although not necessarily to the same point) and the runaway process ends. However, in Lande's models, evolutionary trajectories may lead away from the line of equilibria so that both male trait and female preference can evolve "away from the line at a geometrically increasing rate" (Fig. 3), as Fisher originally claimed. The qualitative conditions for the line of equilibria to be unstable are that females show much heritable variation in mate preference and that there is weak natural selection on the male character. Lande points out that his models help to explain why males of closely related species may differ in a "substantially nonadaptive and random pattern" while the females are phenotypically similar, and also that "random genetic drift may be an important factor promoting speciation by sexual isolation and the evolution of sexual dimorphism".

The similar conclusions reached by models with such dissimilar genetic assumptions underscores the generality of the conclusions. Lande's model, in particular, is general in terms of its assumptions yet clear predictions follow from it[9]. One other idea based on females choosing bizarre males precisely because the male traits constitute handicaps to survival[10] has gained questionable credence among some ethologists who seem not to understand that population geneticists have been unable to model the process successfully. O'Donald[5] presents a detailed analysis of the topic.

The model first sketched by Fisher 67 years ago[11] is only now becoming understood and its widespread implications appreciated. There is no reason why the runaway process should be restricted to plumage characters — vocalizations or pheromone production are other possible candidates. Shortly we might expect relevant parameters to be measured and the efficacy of the theory to be tested. In any event, the finding that, under a variety of circumstances, sexual selection can result in a decline of male viability and thus contribute to population extinction is an important one for evolutionary biology. □

1. Darwin, C. *The Descent of Man and Selection in Relation to Sex* (Murray, London, 1871).
2. Fisher, R.A. *The Genetical Theory of Natural Selection* (Clarendon, Oxford, 1930).
3. Trivers, R.L. in *Sexual Selection and the Descent of Man* (ed. Campbell, B.) (Aldine, Chicago, 1972).
4. Halliday, T. in *Behavioural Ecology, an Evolutionary Approach* (eds Krebs, J.R. & Davies, N.B.) (Blackwell, Oxford, 1978).
5. O'Donald, P. *Genetic Models of Sexual Selection* (Cambridge University Press, 1980).
6. Kirkpatrick, M. *Evolution* **36**, 1 (1982).
7. Lande, R. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3721 (1981).
8. Lande, R. *Genet. Res. Camb.* **26**, 221 (1976).
9. Arnold, S.J. in *Mate Choice* (ed. Bateson, P.P.G.) (Cambridge University Press, 1982).
10. Zahavi, A. *J. theor. Biol.* **53**, 205 (1975).
11. Fisher, R.A. *Eugenics Rev.* **7**, 184 (1915).

'Style' in prehistoric artefacts

from Paul Mellars

VARIATIONS in the forms of hunting weapons have always held a peculiar fascination for archaeologists. In North America, for example, the whole taxonomy of the 'Palaeoindian' cultures depends heavily on the elaborate patterning in the various forms of stemmed and fluted projectile points, whilst in Europe the same preoccupation can be seen in the classification of a wide range of industries from the Upper Palaeolithic period to the Bronze Age.

The tacit assumption underlying these studies has always been that the patterning represented in the projectile point forms reflected in some way the social groupings of the human communities themselves — that the archaeological 'cultures' bore at least some resemblance to the kind of ethnic or cultural groupings which a modern ethnographer or anthropologist might recognize. From the archaeological evidence alone, of course, any correlations of this kind must remain essentially a matter of faith, and it is only with the recent surge of interest in 'ethno-archaeological' studies that the prospect of testing these correlations has emerged.

By far the most important study to date was reported by Polly Wiessner, of the Max Planck Institut für Humanethnologie, at the recent annual conference* of the Prehistoric Society. For several years Wiessner worked among a number of bushman groups (collectively known as the 'San') who still live predominantly by hunting and gathering in the northern and central parts of the Kalahari Desert. Her main aim was to investigate 'style' as expressed in the material culture of these groups, focusing particularly on how far stylistic variation in the different types of artefacts could be correlated with the various linguistic divisions and other forms of social groupings of the present bushman communities. The analysis was conducted at several levels, ranging from variation between individual craftsmen, through the successively larger 'band' and 'band-cluster' groupings, to the level of the three major linguistic divisions (the !kung, G/wi and !xo) which represent the most discrete and ancient social divisions within the societies studied.

From an archaeological perspective, the most instructive results were obtained from a detailed study of the metal-tipped hunting arrows, which are not only among the most highly valued possessions of the bushmen, but are also still used to secure much of the daily food supply. Most important was the discovery that whilst the forms of these arrows showed a good deal of variation from one hunter to another (and sometimes between different batches of arrows manufactured by the same hunter), it was only at the level of the major linguistic grouping that any consistent and statistically significant patterns of variation could be observed. Curiously, variation at this level was apparent not only in relatively minor details of the shapes of the arrow points, (for example, the convexity of the sides or the precise form of the barbs) but also in features that might be expected to have a significant effect on the function of the arrows. Thus, the arrowheads manufactured by the !kung were consistently smaller than those made by the other two groups, whilst those produced by the !xo had distinctively broader, flatter tips. In these features there was no detectable overlap in the forms of the arrows manufactured by the three groups, so individual arrows could easily be visually assigned to one group or another. Despite these variations, the different forms of arrows were used to kill almost exactly the same species of game in all the societies examined.

Most archaeologists would probably argue that these results conform exactly to expectations — that the element of style in the form of the arrow points was serving to differentiate clearly separate social groupings which would have meaning not only to an anthropologist but also to the bushman groups themselves. Wiessner clearly regards this as one of her most important findings, and has referred to this form of patterning as 'emblemic style', that is, the use of stylistic variation as a symbol or 'emblem' which serves to link the individual with some specific, clearly defined social group.

One of the most intriguing implications of her study, however, is that in certain other spheres of material culture, style can operate in a totally different way. The clearest illustration of this was provided by her analysis of the elaborate beadwork used by bushman women to decorate clothing and other personal possessions. In contrast to the clear patterning observed in the arrow-point forms, the patterns of beadwork showed surprisingly little clear-cut variation from one bushman group to another, and instead displayed a bewildering array of variation in the products of the individual women within each group. In other words, style in this case was focused much more at the personal level, than at the level of the local or regional group. Wiessner has therefore referred to this form of highly individualistic style as 'expressive style'.

Wiessner suggests that much of the contrast between the two forms of style can be explained in terms of the particular functions which the different categories of artefacts served in bushman life. The items decorated with beadwork were largely personal possessions which rarely changed hands between individuals and which appeared to play little part in the social relationships maintained between different bushman groups. Arrows, on the other hand, had a much more dynamic role in social life. Not only were arrows frequently exchanged between individual hunters, but the forms of the arrows could potentially be used as a guide to the ownership of killed or wounded game. As Wiessner points out, the ownership of killed game has virtually no significance at the level of the individual hunter, since the strongly egalitarian ethic of bushman society demands that meat from any animal should be widely distributed between members of the local residential group. Even at the level of the separate bands or band-clusters, the question of ownership is of limited relevance, since most of the bands within each major linguistic group maintain a close alliance in economic matters which allows relatively free access to the territories (and therefore food resources) of other groups, and which frequently leads to the movement of individual hunters from one band to another. At these levels of organization, therefore, there is little need for any clear identification of ownership of killed animals, and indeed the uniformity in the form of the arrows could be viewed as an important way of symbolizing and reinforcing the economic and social ties between the different groups. At the level of the linguistic groups, however, social divisions are much more sharply defined. Social or economic exchanges taking place across these boundaries are generally more formalized, and more dependent on a clear recognition of the rights and affiliations of the groups or individuals involved. Wiessner suggests that the clear differentiation in the arrow forms at this level could be important not only economically (by clarifying the ownership of wounded animals) but also in identifying the sources of any aggression or personal conflict across group boundaries. The clear messages encoded in the forms of the arrows thus become an important element in the patterns of 'boundary maintenance' between the societies involved.

Clearly, the archaeological implications of this study need to be treated with caution. Bushman societies have undergone radical transformations in recent years and it is inevitable that many aspects of both bushman life and bushman technology are still adjusting to these changes. Nevertheless, there are several lessons implicit in the bushman data to which the archaeologist should take heed. As a number of authors have recently emphasized (for example, Wobst in *Univ. Michigan Mus. Anthrop., Anthrop. Pap.* **61**, 1977; and Hodder in *The Spatial Organization of Culture*, Duckworth, 1978), 'style' is not a simple univariate

*A conference on 'The Archaeology of Hunter-Gatherers' was held in London on 20–21 March 1982.

P.A. Mellars is in the Department of Archaeology, University of Cambridge, Downing Street, Cambridge CB2 3DZ.

0028-0836/82/240535-02$01.00

phenomenon which can be equated in a simple, one-to-one way with the extent of 'interaction' between separate groups or individuals; it can clearly operate very differently in different social contexts, and can lead to a variety of contrasting and interacting patterns in different spheres of material culture. Variations in the forms of hunting weapons may after all provide one of the most sensitive indicators of group affiliation among prehistoric hunting and gathering communities, but to assume that this would necessarily be true in all social contexts would be stretching the evidence too far. Only by clearly understanding the social mechanisms by which these variations in material culture come into being can we reliably interpret the data from the surviving archaeological record. It is in this context, of course, that the ethnoarchaeological studies carried out by Wiessner and others are of central importance to the understanding of human behaviour in the past. □

Total eclipse of June 1983

from A.D. Fiala

NEXT June 11, nature will provide one of her most spectacular sights, a total eclipse of the Sun. The eclipse will be of long duration and visible from accessible land sites where there is every prospect of good weather — a favourable combination that has not occurred since June 1973 and will not occur again until July 1991. The path of totality will cross land mostly in Indonesia, but also in Papua New Guinea and a few small islands. Maximum duration of totality will reach over 5 minutes 15 seconds, at a point midway between the islands of Java and Sulawesi. However, on both islands duration in excess of five minutes will occur, with the Sun well up in the sky.

The path of totality will start in the south Indian Ocean at sunrise. For the first hour the umbra will sweep over water, then fall upon Christmas Island. It will then travel the Indonesian archipelago, crossing the centre of Java, the south of Sulawesi and the south of New Guinea, with Ujung Pandang and Port Moresby lying nearly on the central line (see the map). On Java, several large cities lie in the path, including Jogjakarta, Surabaya and Semarang. Bosscha Observatory, the closest to the equator of the world's major observatories, will lie outside the path of totality, however, and will see only 97 per cent partial phase.

The weather prospects for the eclipse are excellent, according to Canadian meteorologist Jay Anderson (*Newsletter of the Bull. R. astr. Soc. Can.*, June 1981). Monsoon conditions give a probability of clear sky of over 65 per cent on the north-east coast of Java and south-west coast of Sulawesi, and nearly 90 per cent at Port Moresby. The sunspot cycle will be midway between maximum and minimum, so intermediate coronal structure and activity may be expected. The eclipse will last long enough for the sky to darken, making it easier to see faint features, particularly from high altitude.

Many astronomers and physicists are laying provisional plans as, indeed, is the Indonesian Government, which is hoping to attract tourists to both the eclipse and the hundredth anniversary of the eruption of Krakatoa that falls at the same time. Experiments are being planned to measure the size of the Sun, to probe the composition and structure of the solar atmosphere and to look for solar effects on the Earth's atmosphere and on the behaviour of living organisms.

A collaborative team from the US Naval Observatory, the Computer Science Corporation, NASA Goddard Space Flight Center and the Astronomical Society of Canberra plans to man at least one station near each edge of the path, in order to measure the solar diameter. This procedure has been carried out at many eclipses, since the original measurements in 1715 by Halley. So far, the results indicate that there is a variation in the size of the Sun, but too few measurements have been gathered to distinguish between monotonic and pulsating variation. The long duration of this eclipse should give the opportunity for observing an extraordinarily large number of small limb features. The observations are not at all difficult, requiring only careful timing and concentration.

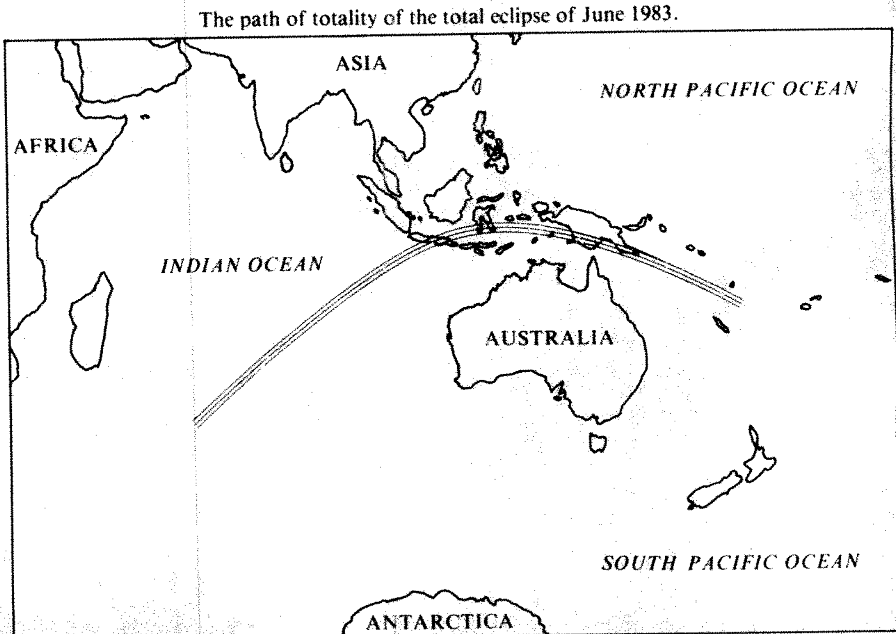
Nowadays at any major eclipse there are several airborne observers. NASA's Kuiper Airborne Observatory may be flying along the track over the Indian Ocean, using the Moon as a high-resolution image dissector along the limb to study the limb brightness variation in sub-millimetre wavelengths.

On the ground, several US experimenters hope to be present. A team led by D. Landman, Institute of Astronomy, Hawaii, hopes to continue spectrophotometry of the weak emission lines and continua of ordinary prominences to provide tests of theoretical models. Investigative teams from Kitt Peak, Sacramento Peak and Williams College would like to study in different ways the motion of material in the inner corona, while a group from Iowa State University would study the motion of dust in the outer corona.

Turning the Sun 'off' for several minutes should produce some significant effects in the Earth's atmosphere in the wake of the eclipse path. At least four teams from the US are hoping to study these effects. Two of them would attempt to refine our knowledge of the atmospheric turbulence causing shadow bands and another would look for effects of the wake as far away as India. Observing teams are also expected from Japan and Australia, and quite likely from India, Europe and Canada as well. Indonesia has created a national committee on the eclipse.

This eclipse is the first of three total eclipses to occur in the region over a five-year period. Interestingly, the second is also the next total eclipse, on 22–23 November 1984, and its path intersects with the 1983 eclipse in the Gulf of Papua. However, no experiment has yet been proposed which would make use of this circumstance. The third eclipse will be on 18 March 1988 over Sumatra and Borneo.

The path of totality of the total eclipse of June 1983.



A.D. Fiala is at the US Naval Observatory, 34th and Massachusetts Avenue, NW Washington, DC 20390.

0028-0836/82/240536-01$01.00

Ionospheric hole caused by rocket engine

from Michael J. Rycroft

AMONG the most interesting of the wide variety of active experiments[1] that can be carried out to study the tenuous upper atmosphere and ionosphere are those that are 'accidental'. One such experiment, perhaps better termed an 'experiment of opportunity', is the operation of a rocket engine at heights above 200 km. The Saturn V launching of Skylab on 14 May 1973, which halved the total electron content of the ionosphere for three hours[2,3], created considerable interest in the effects of rocket exhaust gases on the ionosphere. Further measurements of an artificially induced 'ionospheric hole' were made during the Atlas-Centaur launch on 20 September 1979 of HEAO-C[3,4]. Michael Mendillo and Jeffrey Baumgardner, of Boston University, report[5] the first results of the latest experiment of this type, that for the Atlas-F launch of the NOAA-C weather satellite on 23 June 1981. Launched from Vandenburg Air Force Base, California, at around 03.00 local time, the rocket engine burned to an altitude of 434 km, well into the topside F-region of the ionosphere.

Because the rocket exhaust gases contain highly reactive molecules such as H_2, H_2O and CO_2, reactions with atomic oxygen ions (which are dominant in the F-region) produce molecular ions such as H_2O^+ and H_3O^+. The rates of the reactions are between a hundred and a thousand times faster than those which normally occur with ambient atmospheric nitrogen and oxygen molecules. Thus the atomic ion plasma is rapidly converted to a molecular ion plasma which then recombines quickly with ionospheric electrons, removing ionization and creating the local ionospheric hole. As the rocket proceeds and the plume of exhaust gases diffuses through the atmosphere, the electron content of a column through the F-region, both bottomside and topside, is depleted. The dissociative recombination reactions of the molecular ion plasma, which create the hole, produce oxygen atoms in the singlet D excited state. As these fall into triplet P states, red airglow is emitted, primarily at a wavelength of 630 nm. The position where the airglow is emitted is determined by the time scale of the diffusion of rocket exhaust species through the atmosphere.

The two important quantities to be observed in such experiments are the airglow and the total electron content (TEC) of the ionosphere. In the June 1981 experiment, the TEC was found by measuring the Faraday rotation of the plane of polarization of 136 MHz radio beacon signals from the ATS-1 geostationary satellite over the Pacific Ocean, the groundsite for such observations being chosen such that the ray path passed through the rocket's exhaust plume at 350 km altitude. Four minutes after launch, the TEC decreased rapidly, and reached a steady value some three minutes later; the TEC was found to decrease by 1.7×10^{17} electrons per m^2. From the same site narrow-band optical observations were made using a low-light level image-intensified photographic system with a 60° field of view. As the TEC decreased, an expanding shell of airglow with an intensity of several kiloRayleighs was observed; in the horizontal plane, this appeared rather like a 'smoke ring' with a radius of up to about 1,200 km. As the shell expanded more, its intensity decreased on a time scale of several minutes.

What can be deduced from such observations of the expanding shell of rocket exhaust molecules? First, the diffusion coefficient can be estimated to be $3 \times 10^7 m^2 s^{-1}$. This is consistent with the lightest species of exhaust gases (H_2) diffusing through an ambient atomic oxygen atmosphere, with a temperature of 1,000K at 350 km altitude. Second, knowing that the Atlas-F rocket ejects 10^{27} H_2O molecules per s (about 50 kg s^{-1}), the observation that the airglow emission was maximal two minutes after the initiation of the hole can be interpreted. It is found that 14 per cent of the plasma recombinations yield oxygen atoms excited in the singlet D state. Also, it is shown that, to first order, quenching of the excited oxygen atoms by molecular species is not important.

Further analysis of this and similar experiments can be expected in the future, and other chemical release experiments are planned, for example from the Space Shuttle. Even the operation of the Shuttle's engines or the dumping of water produced by its fuel cells could lead to interesting results. Such active space plasma experiments conducted in the ionosphere are not only of scientific interest, but also have practical significance for various aspects of radio communications. □

1. Rycroft, M.J. *Nature* **287**, 7 (1980).
2. Mendillo, M., Hawkins, G.S. & Klobuchar, J.A. *J. geophys. Res.* **80**, 2217 (1975).
3. Mendillo, M. *Adv. Space Res.* **1**, No.2, 275 (1981).
4. Mendillo, M., Rote, D. & Bernhardt, P. *EOS Trans. Am. geophys. Un.* **61**, 529 (1980).
5. Mendillo, M. & Baumgardner, J. *Geophys. Res. Lett.* **9**, 215 (1982).

Michael J. Rycroft is Head of the Atmospheric Sciences Division of the British Antarctic Survey, NERC, Madingley Road, Cambridge CB3 0ET.

Coastal plants take to the road

from Peter D. Moore

IT has long been recognized that the spreading of salt on roads in Britain during the winter is having a progressive influence on the vegetation of our roadside verges[1]. A similar problem has been described from the United States back in 1969, when it was estimated that 6×10^6 tons of salt were applied annually to roads in the northern states[2]. Up to 10^6 tons are used annually in Britain[3], and locally this often means that salt is being added to the soil of these verges at a rate of 3–4 kg m^{-2} yr^{-1} (in the United States, values of up to 6 kg m^{-2} yr^{-1} have been recorded[2]).

If we compare these figures with those obtained from studies of natural salt spray on coastal vegetation, such as maritime cliff grasslands, then the ecological impact of the salt input can be appreciated more fully. In his studies of the Lizard peninsula, Cornwall, Malloch[4] determined the annual 'salt' (Na, K, Mg, Ca, Cl) deposition in rain gauges to be approximately 0.07 kg m^{-2} yr^{-1}. In only one location did he record a value as high as 0.86 kg m^{-2} yr^{-1} of sodium, which is approximately equivalent to 2.6 kg m^{-2} yr^{-1} of total salt. The results are in keeping with those from sites in western Ireland 15 km from the sea[5] (0.04 kg m^{-2} yr^{-1}), in Lancashire 2 km from the sea[6] (0.01 kg m^{-2} yr^{-1}) and in Cheshire 40 km from the sea[6] (0.004 kg m^{-2} yr^{-1}). It is evident, therefore, that our roadside verges are receiving a considerably greater (often by a factor of 50) input of salt than our most exposed cliff grasslands, and it is not surprising that the effects of such treatment are being increasingly felt.

The application of salt to soil has a strong influence on its structure, and leads to clay flocculation and consequent loss of aeration. It also, naturally, raises the osmotic potential of the soil water, making it difficult for plants which lack physiological adaptation to absorb water. There are two main consequences for the vegetation: first, there will be a strong selective pressure against those individuals within populations that lack the requisite adaptations; and second, halophytic, maritime plant species will be able to invade these inland habitats.

Relatively little data are available on the effects of salt application upon the general

Peter Moore is in the Department of Plant Sciences, University of London King's College, 68 Half Moon Lane, London SE24 9JF.

0028-0836/82/240537-02$01.00

balance of grassland species but it seems that white clover (*Trifolium repens*) and meadow grass (*Poa pratensis*) may suffer more than red fescue (*Festuca rubra*), rye grass (*Lolium perene*) and crested dogstail grass (*Cynosurus cristatus*)[7]. Selection of specifically adapted genotypes within populations is difficult to demonstrate in the field, but experimental studies suggest that this will be an active process[8].

The second conseqùence — the invasion of typically coastal species of plant — is more obvious to the observer and hence has been studied in more detail. In 1976, Mathews and Davison[9] surveyed the Newcastle-upon-Tyne area and found several such species (such as *Aster tripolium, Puccinellia distans, Plantago maritima* and *Sueda maritima*) on roadside verges as much as 13 km from the sea. The salt marsh grass (*Puccinellia distans*) has now been recorded on verges extensively throughout northeastern England, in the Midlands, East Anglia and Kent[10]. Other species have not kept up with the rapid spread of *Puccinellia,* though the sea plantain (*Plantago maritima*) and the sea spurrey (*Spergularia marina*) are still advancing gradually. The sea aster (*Aster tripolium*), however, appears to have failed in its bid for a place in the new saline grasslands; seed set has been poor and populations are contracting rather than expanding. Seed dispersal techniques could be the key to success in this invasion. *Puccinellia, Plantago* and *Spergularia* are low-growing plants and their seeds could easily be carried in the mud adhering to vehicle tyres, while *Aster* seeds are produced well above the ground and rely on wind dispersal.

A survey of roadside vegetation in Norfolk has also recently been conducted by Bull[11]. Permanent recording plots were set up five years ago and the species present recorded annually since then. Both extinctions and immigrations have been observed during this time and Bull regards many of the new immigrants (for example, *Polygonum aviculare, Artemisia vulgaris* and *Rumex obtusifolius*) as salt tolerant. The difficulty in interpreting this kind of data is the number of variables involved (frequency and time of mowing in particular); controlled experimental studies are evidently needed.

The invasion of our roadside verges by halophytes is thus proceeding in much the way that the 'railway flora' developed during Victorian times[12]. One difference is that the nature of the current change will be less conspicuous. Extinctions may be more abundant than immigrations, but these are less easily noticed by casual observers, and genetic modifications within our roadside plants may also be overlooked. We may be sure, however, that there are subtle changes afoot in our 1,500 km² of roadside verges[13]. □

1. Moore, P.D. *Nature* **263**, 189 (1976).
2. Westing, A.H. *Phytopathology* **59**, 1174 (1969).
3. Davison, A.W. *J. appl. Ecol.* **8**, 555 (1971).
4. Malloch, A.J.C. *J. Ecol.* **60**, 103 (1972).
5. Sparling, J.H. *J. Ecol.* **55**, 1 (1967).
6. Allen, S.E., Carlisle, A., White, E.J. & Evans, C.C. *J. Ecol.* **56**, 479 (1968).
7. Bannister, P. *Introduction to Physiological Plant Ecology*, 198 (Blackwell Scientific, Oxford, 1976).
8. Rush, D.W. & Epstein, E. *Pl. Physiol.* **57**, 162 (1976).
9. Mathews, P. & Davison, A.W. *Watsonia* **11**, 146 (1976).
10. Scott, N.E. & Davison, A.W. *Watsonia* **14**, 41 (1982).
11. Bull, A. *Trans. Norfolk Norwich Nat. Soc.* **26**, 23 (1982).
12. Lousley, J.E. in *The Flora of a Changing Britain* (ed. Perring, F.) 73 (Classey, Hampton, 1970).
13. *Amenity Grasslands, the Need for Research* (NERC report, 1977).

Seawater under extreme pressure

from Michael Whitfield

UNTIL recently, the experimental difficulties involved in performing precise measurements at high pressures have ensured that relatively little information was available about the properties of seawater and the reactions of its constituents in ocean conditions. The available information has been scattered throughout the literature, uncorrelated and accessible only to the *cogniscenti.* Two recent publications[1,2] indicate that this picture is changing rapidly.

Most experiments are carried out at atmospheric temperature and pressure, but it is a harsh reality that mean ocean conditions lie closer to 4°C and 200 atm, and more extreme conditions are not uncommon. The pressure generated by the overburden of water has a significant influence on the density of seawater and hence on the stability of the water column and the direction and rate of water movement. Furthermore, minerals such as calcium carbonate and silica, generated in surface waters by biological processes, become progressively more soluble as they settle through the ocean, encountering waters at increasing pressures. Consequently the recycling of these minerals and the many trace elements associated with them is strongly dependent on the pressure sensitivity of the dissolution reactions. The solubility of calcium carbonate in the deep oceans is of particular concern since the dissolution of this mineral is likely to provide the final sink for the excess carbon dioxide currently being injected into the atmosphere by the combustion of fossil fuels. Another reason for growing interest in these measurements is the discovery of rich and diverse communities of organisms surrounding warm hydrothermal oases along the mid-oceanic ridges which are subjected to pressures of several hundred atmospheres.

To appreciate the significance of the recent developments one must look first at the problem of defining the equation of state for seawater, which enables the density of seawater to be calculated at a particular temperature, pressure and salinity. These parameters can be accurately measured and the calculated density fields then used to elucidate oceanic circulation patterns.

To establish the new equation of state for seawater, techniques were developed for the precise measurement of solution density over a wide range of temperature and pressure. Density measurements at atmospheric pressure for solutions of individual electrolytes found in seawater can be used to estimate the contributions made by the component ions to the solution volume (the ionic molal volumes). Similar measurements made at high pressure provide estimates of the ionic molal compressibility. The influence of pressure on a chemical reaction can be estimated from the accompanying volume change — a reaction which produces a decrease in volume being favoured by the application of pressure.

Millero[2] has taken advantage of recent density measurements at atmospheric pressure and at high pressure to compile data on ionic molal volumes and compressibilities over the temperature and pressure range of interest to oceanographers (0–50°C, 1–1,000 atm). Using this information he has calculated the effect of pressure on the solubility of a range of mineral phases (including silica, strontium carbonate and calcium carbonate) and shown that the calculated values are in good agreement with available experimental data. Furthermore, by taking into account the molal compressibilities he was able to show that, for the minerals he considered, there was no need to invoke the presence of structurally altered zones at the surface of the solid phase to explain the observed pressure dependence. However, molal volume measurements over a wider temperature and concentration range will be needed in order to estimate the effects of pressure on reactions in the geothermal brines produced along saline spreading centres on the ocean floor. It is now becoming apparent that the reactions occurring at these sites may be responsible for fixing the concentration of many seawater components, and that the more spectacular deep-ocean vents may be surrounded by rich mineral deposits whose formation and stability require further investigation. □

Michael Whitfield is at the Marine Biological Association of the United Kingdom, Plymouth PL1 2PB.

1. UNESCO *Background Papers and Supporting Data on the International Equation of State for Sea Water 1980* (Paris, 1981).
2. Millero, F.J. *Geochim. cosmochim. Acta* **46**, 11 (1982).

0028-0836/82/240538-02$01.00

The impact, transmission and evolution of infectious diseases

from Robert M. May

OVER the past ten years and more, the World Health Organization (WHO) has increasingly placed its emphasis in the Third World on primary health care — paradigmatically exemplified by 'barefoot doctors' in villages — rather than on heroic technology and large hospitals[1]. I think this is very sensible, as many of the afflictions that combine to make life spans in the Third World markedly shorter than those in the developed countries can be countered by relatively inexpensive and simple measures; rehydration treatments for infants with diarrhoea or public health practices that help lower parasite burdens give better value for each dollar expended than do CAT scanners.

In designing programmes of primary health care, insights into the overall transmission and maintenance of infections within host populations are valuable. This fact has paradoxical aspects, for to some the 'flute music' (Simon Levin's phrase) of mathematical epidemiology must seem like the most baroque manifestation of Western cultural imperialism, even less relevant to village paramedics than is open heart surgery. Motivated partly by the idea that theoreticians and practitioners have interesting things to say to each other in this area, a recent Dahlem Conference* brought together ecologists, population biologists, epidemiologists, veterinarians and a variety of medical people to examine the overall population biology of infectious diseases. Taking several background papers as read, the conferees divided into four groups to focus on the impact of diseases on human and other animal populations; the factors influencing transmission and maintenance of infection; the possibilities for control or eradication of infections; and the broad co-evolution of hosts and pathogens.

Impact

As reviewed for the meeting by Holmes (University of Alberta), anecdotes abound about the devastation that diseases can inflict on natural populations of plants and animals. It is, however, much more difficult to assess the extent to which diseases are important in regulating such populations. For instance, in some species of wildfowl in North America some 80–90 per cent of the individuals not shot by hunters die of diseases each year, yet it is arguable that availability of breeding sites remains the primary factor regulating population density. The discussion made it plain that the various causes of deaths are hard to disentangle in most vertebrate populations, and left the impression that regulatory mechanisms may often depend on the interplay of several factors. Thus in those parts of Europe where rabies is now endemic in fox populations, the fox density may be significantly depressed below the disease-free level, which is set by territorial considerations; the present densities appear to depend on both the disease and territory size.

Holmes advanced the suggestion that invertebrate populations may more typically be regulated by infectious diseases and parasites than are vertebrate ones. Although some empirical evidence points in this direction, a systematic study is lacking. Theory suggests[2] that host populations are more easily regulated by infectious diseases if there is little or no acquired immunity (typically making the regulation of invertebrates, which lack such acquired immunity, easier than of vertebrates), and if the intrinsic rate of growth of the host population is not too large.

In any assessment of the impact of disease on natural populations, *Homo sapiens* must be considered as a special case. This is not out of respect for Bishop Wilberforce and his spiritual heirs, but because the patterns of explosive growth of human populations over the past century and more are without parallel elsewhere in the animal kingdom.

Changing life expectancies in human populations, particularly over the past two centuries in the developed world and since World War II in the developing world, are almost wholly due to reduced mortality from infectious diseases[3,4]. Although these diminished mortality rates and their demographic consequences are clear, the relative importance of the contributions made to them by better nutrition, improved hygiene, medical advances and other factors remains the subject of sharp and unresolved debate[3].

Transmission

Many biologists and medical people are fascinated by the intricate details that make each association between a host species and a virus, bacterium, protozoan or helminth unique. But underlying this richness and peculiarity of detail — which must be kept in sight as one moves towards control programmes — are some common themes.

A lot of discussion centred around the extent to which R_0, the 'intrinsic reproductive rate' of the parasite[5-8], can be used to characterize particular host–parasite associations. Calculation of R_0 itself involves classifying parasitic infections according to their overall dynamics rather than their conventional taxonomy. For microparasites, which broadly are those with high rates of direct reproduction in the definitive host (as typified by most viral and bacterial, and many protozoan, infections), R_0 is the average number of secondary infections produced when one infected host is introduced into an uninfected population. For macroparasites, which broadly are characterized by having no direct reproduction within the definitive host (as is typified by most helminth infections), R_0 is the average number of adult offspring established in 'second generation' hosts that one adult parasite produces in an essentially uninfected host population. In either case, R_0 depends both on the basic biology of the host–parasite interaction, and on the environmental and social factors that influence transmission rates.

An accurate estimate of R_0 for microparasites can be made if good serological studies are available. Failing this, rough estimates can be made from the approximate formula[6] $R_0 = 1 + (L/A)$ where L is the average life span of the human or other host, and A the average age at which infections are acquired when the disease is endemic. Estimation of R_0 for macroparasites is more difficult, and involves knowledge of the average parasite burdens, of the way in which parasites are distributed within the host population (almost invariably in a clumped or 'overdispersed' manner, rather than independently, randomly or uniformly), and of the way egg output and adult parasite survival depend on the number of parasites in a host.

In addition to attempting to estimate R_0 for a selection of microparasites and macroparasites, the conferees also catalogued some of the difficulties inherent in the concept. For one thing, not all infections lend themselves to ready classification in this binary scheme. For another, some microparasites have two or more qualitatively distinct categories of host. For example, while typhoid usually runs a relatively short course of infection from which the host either recovers or dies, a few hosts ('Typhoid Marys') will be asymptomatic yet will carry and can transmit the infection for a relatively long time, possibly even lifelong; it is not obvious which is more important in transmission, the much larger number of much shorter infections, or the small number of long and asymptomatic infections. Gonorrhoea and some other infections may be primarily maintained by a core of 'superspreaders'[9-12]. Such refinements can be embraced within the definition of R_0, but they complicate its estimation.

Control

Control is a term often used loosely to describe any reduction in the incidence of an infection. Clearly, eradication of the infection is a limiting case. Whatever the

*The Dahlem Conference on the 'Population Biology of Infectious Diseases' was held in Berlin on 15–19 March 1982. The proceedings will be published as a book by Springer Verlag, later this year.

Robert M. May is Class of 1877 Professor of Zoology at Princeton University, Princeton, New Jersey 08544.

0028-0836/82/240539-02$01.00

degree of control being attempted, the goal will tend to be relatively easier if R_0 is small. It was remarked by Molineaux (WHO, Geneva) that the WHO success in the global eradication of smallpox and the failure in efforts to eradicate malaria from large parts of Africa may be partly because for smallpox in developing countries R_0 is $\sim$ 3 or 4, while for *falciparum* malaria R_0 ranges from $\sim$ 50 to several hundreds.

The existence of reasonably safe and effective vaccines against many microparasitic infections provides the basis for attempts at control or even eradication. Again, knowledge of R_0 can give a rough idea of the prospects for such campaigns. Assuming the vaccine to be 100 per cent effective (which few are), eradication requires that the fraction, p, of the population protected by vaccination must exceed a critical value[6,7,13] $p > 1 - 1/R_0$. For large R_0, this can require a forbiddingly high proportion of the population to be vaccinated. Thus, although the aetiologies of measles and smallpox are in many ways similar, with the infection always being apparent and running a relatively short course, the tentative estimate that R_0 is probably around 10–20 for measles in developing countries may make its global eradication much more difficult than was the case for smallpox with its R_0 of 3–4.

For macroparasites, there was discussion of the distinction between infection (having one or more parasite) and disease (having a parasite burden sufficiently high to produce illness). Especially if parasite burdens are highly overdispersed, as they usually are, relatively few hosts may carry burdens high enough to induce illness, even though essentially everyone is infected. There can thus, as emphasized by Warren[14] and others, be a real distinction between 'infection' and 'disease', and it may make sense to target control measures against the heavily infected hosts.

The concept of a 'breakpoint', first introduced by Macdonald[15] in his thinking about control of schistosomiasis, is an interesting one for many macroparasitic infections. If the parasite has a sexual stage in the definitive host, as many macroparasites do, then there can be two alternative stable states for the host–parasite association above the threshold at $R_0 = 1$: if the parasite population is present at densities above the 'breakpoint' level, the chances that female parasites in a given host will be mated is high, and the parasite population will be able to realise its intrinsic reproductive rate and establish itself; but if the parasite population is initially at a density below the breakpoint level, the chance for a given female to acquire a mate is low, and the infection will die out (even though R_0 for mated females exceeds unity). This notion of a breakpoint is obviously attractive. It offers the seductive hope that macroparasites could be eradicated by a one-time reduction below breakpoint densities, without the environmental or social changes that are needed if R_0 itself is to be reduced below unity. This contrasts with most other parasites, where, in the absence of a breakpoint phenomenon, unremitting vigilance is necessary for parasites with high R_0, even after the infection has been eradicated from a region; the basic biology is such that the infection will tend to re-establish itself if a few infected individuals are introduced. Reviewing recent work on this subject[16,17], the conferees sadly concluded that such breakpoint densities are probably too low to be of practical significance (mainly because the patterns of parasite clumping in hosts, mentioned above, tend to make mated pairs likely even at low density).

Throughout the week, it was often observed that many of the infections that excite attention in the West are of little significance in any global league table of diseases. For example, increasing incidence of various infections associated with homosexuals in large cities is attracting considerable attention, yet [as summarized for the conferees in Paul Mann's (Public Health Laboratory, Bath) clerihue] "Malaria/Is scarier/Than catching clap/From a travelling chap".

Evolution

The notion that 'successful' or 'well adapted' parasites are relatively harmless to their hosts is set forth as the received wisdom in most medical texts, and elsewhere. The idea is, at first sight, not unreasonable: all else being equal, it is to the advantage of both host and parasite for the parasite to inflict little damage. Moreover, a certain amount of anecdotal information supports this view. In his background paper, Allison (System Research, Palo Alto) observed that, for example, in regions of Africa where trypanosomiasis is endemic, indigenous ruminants suffer mild infections with insignificant morbidity, while domestic ruminants that have been bred for a long time in the region suffer more severely, with significant morbidity and mortality, and recently imported exotic ruminants suffer virulent infections which are usually fatal if untreated.

From a theoretical point of view, it would indeed appear that parasites evolve to be avirulent, provided that transmissibility and duration of infectiousness are entirely independent of virulence. This assumption, however, is not generally valid; the damage inflicted on their hosts by most viral, bacterial, protozoan and helminth parasites is often directly associated with the production of transmission stages. Once these complications are introduced into the theoretical models[18-20], it appears that many co-evolutionary paths are possible, depending on the details of the interplay between the virulence and the transmissibility of the parasite.

Considerable discussion centred around one unusually well documented example[21,22]. In the early 1950s, the myxoma virus was introduced into rabbit populations in Australia and England. At first the disease was highly virulent, but over the subsequent decade successively less virulent strains of the virus began to appear. Since the mid 1960s, the virus appears to have come to an equilibrium with its rabbit host (in both Australia and England), with the predominant strain of the virus being one of intermediate virulence. These data can be analysed to get an approximate estimate of the relationship between the virulence and the transmissibility of the various strains of the myxoma virus[20]; the nature of this relationship is such that strains with intermediate virulence have the largest R_0, and may thus be expected to predominate.

There was discussion of other cases where the evolutionary pressures on the parasite may make for its being highly virulent. The various baculoviruses, which kill their insect hosts and effectively turn them into masses of viral transmission stages, are one such example that may be amenable to an explicit analysis of the contending evolutionary forces.

The group concentrating on evolutionary aspects of host–parasite interactions thus concluded that, although parasite 'harmlessness' may characterize many old-established associations, neither *a priori* arguments nor empirical evidence points towards its being a general rule. Several interesting questions that seem ripe for mathematical or experimental investigation were identified.

In all, most of the participants found the conference helpful both in summarizing where things stand, and in identifying directions in which we might productively move. Bryan Clarke spoke for everyone when he said:

If you're teased by a sense of unease,
We hope that this volume will please,
That the work clears the murk
From the quirks that may lurk
In the laws of the cause of disease.

1. Mahler, H. *Scient. Am.* **243** (Sept.), 66–77 (1980).
2. Anderson, R.M. & May, R.M. *Nature* **280**, 361, 455 (1979).
3. McKeown, T. *The Role of Medice: Dream, Mirage or Nemesis* (Princeton University Press, 1979).
4. Wrigley, E.A. & Schofield, R.S. *The Population History of England, 1541–1871* (Harvard University Press, 1981).
5. Macdonald, G. *Trop. Dis. Bull.* **49**, 813 (1952).
6. Dietz, K. in *Epidemiology* (eds Ludwig, D. & Cooke, K.L.) 104 (SIAM, Philadelphia, 1975).
7. Anderson, R.M. & May, R.M. *Science* **215**, 1053 (1982).
8. Yorke, J.A., Nathanson, N., Pianigiani, G. & Martin, J. *Am. J. Epidemiol.* **109**, 103 (1979).
9. Yorke, J.A., Hethcote, H.W. & Nold, A. *J. Sex Trans. Dis.* **5**, 51 (1978).
10. Nold, A. *Math. Biosci.* **52**, 227 (1980).
11. Kemper, J.T. *Math. Biosci.* **52**, 227 (1980).
12. May, R.M. *Nature, News and Views* **291**, 376 (1981).
13. Smith, C.E.G. *Proc. R. Soc. Med.* **63**, 1181 (1970).
14. Warren, K.S. *A. Rev. pub. Hlth* **2**, 101 (1981).
15. Macdonald, G. *Trans. R. Soc. trop. Med. Hyg.* **59**, 489 (1965).
16. May, R.M. *Math. Biosci.* **35**, 301 (1977).
17. Bradley, D.J. & May, R.M. *Trans. R. Soc. trop. Med. Hyg.* **72**, 262 (1978).
18. Levin, S.A. & Pimentel, D. *Am. Nat.* **117**, 308 (1981).
19. Bremermann, H.J. *J. theor. Biol.* **87**, 671 (1980).
20. Anderson, R.M. & May, R.M. *Parasitology* (in the press).
21. Fenner, F. & Radcliffe, F.N. *Myxomatosis* (Cambridge University Press, 1966).
22. Ross, J. in *Animal Disease in Relation to Animal Conservation* (Symposium Zoological Society, London, in the press).

REVIEW ARTICLE

The relationships of *Sivapithecus* and *Ramapithecus* and the evolution of the orang-utan

Peter Andrews* & J. E. Cronin†‡

* Department of Palaeontology, British Museum (Natural History), Cromwell Road, London SW7 5BD, UK
† Department of Anthropology, Harvard University, Cambridge, Massachusetts 02138, USA

We review here the molecular data that bear on and provide a framework for interpreting hominoid relationships. Man is shown to be most closely related to chimpanzees and gorillas among extant hominoids, with the orang-utan more distantly related to them and the gibbons more distantly still. A fossil ape, Sivapithecus meteai, *shares several characters with the orang-utan and is thus probably related to it.* S. meteai *is part of the Middle Miocene* Sivapithecus–Ramapithecus *species complex, and if this group forms a valid clade then* Ramapithecus *must also be considered as being more closely related to the orang-utan than to man. The date of divergence of the orang-utan from the African apes and man is suggested by fossil and molecular evidence to be 10±3 Myr ago.*

THE Miocene epoch (25–5.5 Myr) was the period that heralded the appearance of mammalian faunas of modern appearance. It was also the period that saw the main radiations of the hominoid primates, culminating in forms that share many of the characteristics of the living apes and man[1,2]. The specific affinities within this complex of fossil and living hominoids are still, however, subjects of much controversy[1–3], and other than recent Pleistocene forms which are obviously related to the living orang-utan[4] and gibbons[5], and the Plio-Pleistocene forms of early hominids from East Africa[6,7], no fossil hominoid has been convincingly shown to be related uniquely to any of the extant hominoids.

Foremost in this controversy is the postulated hominid affinity of *Ramapithecus*, a form found now in middle to late Miocene deposits in East Africa[1,8], Turkey[9], China[10] and Indo-Pakistan[3,11,12]. This phyletic linkage has been much debated and has been both rejected and supported on morphological[11,13] and molecular grounds[14–16]. Central to understanding this problem is: (1) the relationship of the complex of species within the *Sivapithecus–Ramapithecus* network, and (2) the relationships among the living hominoids and in particular the status of the orang-utan. To a large extent the adaptation and relationship of the orang-utan have remained substantial enigmas[14,17,18], but they are clearly pivotal to our understanding of ape and hominoid evolution.

Recently, the relationships among the living species of hominoids have been greatly clarified by macromolecular comparisons[14,15,17,19]. Although not yet definitive, these comparisons have provided a new insight into the branching sequence or cladistic relationships of the living apes and the timing of the divergence between apes and man. This evidence will be reviewed here, with particular reference to the relationships of the orang-utan with the other hominoids. Changes in morphology, particularly of the teeth and face, will be correlated with the pattern of relationships thus established.

As regards the fossils, a new fossil hominoid specimen has been recently described which sheds a substantial amount of light on the *Sivapithecus–Ramapithecus* complex. *S. meteai*, from the late Miocene of Turkey[20], shows similarities in morphology and a possible phyletic relationship to the living orang-utan. *S. meteai* clearly shares broad similarities with other Miocene *Sivapithecus*[20] forms, which in turn are closely allied to *Ramapithecus*[11]. If this linkage is correct then one segment of the hominoid evolutionary tree can be rooted through a purely palaeontological approach.

Molecular affinities among the hominoids

Each organism carries within itself a history that is encoded in the genome. Given the appropriate techniques we can decipher this codex and elucidate phylogenetic history. Such macromolecular comparisons may yield insights into both the sequence and timing of ancient speciation events. The molecular approach has long been critical to the resolution of several controversies concerning phyletic links over a wide taxonomic scale, for example, in the case of taxa such as Aves, Amphibia, Reptilia and Mammalia, particularly Primates. Such comparisons are not limited to living forms, as tissue from a frozen mammoth has supported its close molecular relationship to the modern elephant[21], and skin from the recently extinct Tasmanian wolf[21] has confirmed its affinities with the dasyuroid group of marsupials.

With respect to hominoid evolution, substantial genetic data are now available which shed light on the evolution of humans and apes. Data from immunodiffusion, microcomplement fixation, radioimmunoassay, amino acid sequencing, electrophoresis, nucleic acid hybridization, nucleotide sequence and restriction endonuclease mapping, and cytogenetics, have shown a fairly consistent pattern of the branching sequence or cladistics of the hominoids, and working within this framework one can attempt to analyse the rate and nature of morphological evolution along the specific lineages.

The biomolecular evidence (Fig. 1) argues for three clear major cladogenic events in hominoid evolution: first, the separation of the gibbon lineage from that leading to the great apes and man: second, the subsequent divergence of the orang-utan lineage from that linking the African apes and man; and thirdly, evidence from molecular comparisons that *Homo*, *Pan* (chimpanzee) and *Gorilla* share a substantial lineage before their subsequent divergence. Evidence as to which of the two extant lineages shares a period of common ancestry to the exclusion of the third is at present equivocal; evidence exists which favours pairing of *Pan–Gorilla*, *Homo–Pan* and *Homo–Gorilla* (in order of greatest preference). The difficulty in determining the sequence of divergence is primarily a result of extremely short

‡ Present address: Department of Anthropology, University of California, Berkeley, California 94720, USA.

0028–0836/82/240541—06$01.00

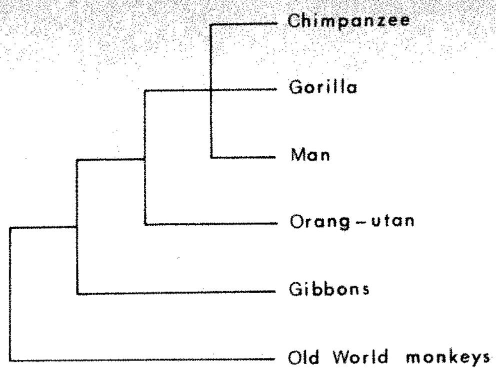


Fig. 1 A molecular cladogram of the Hominoidea incorporating data from immunological studies, electrophoretic analysis, protein sequencing, nucleic acid hybridization and restriction endonuclease analysis of mitochondrial DNA. The data are discussed in the text and in refs 22 and 23 in more detail. Molecular clock calibration of the various sequences of divergence (as calibrated in the text and in refs 15, 22 and 27) are (1) *Homo–Pan–Gorilla*, 5 ± 15 Myr ago; (2) *Pongo–Homo, Pan, Gorilla*, 10 ± 3 Myr ago; and (3) *Hylobates–Pongo, Homo, Pan, Gorilla*, 12 ± 3 Myr ago.

lineages, in a temporal sense, and of the experimental error inherent in the techniques, and so far no data have conclusively demonstrated any unique shared derived molecular changes for any pair of the lineages. However, as interesting as this issue is, it is not central to the evidence for the common ancestry of the African apes and man and to the phyletic position of the orang-utan (Fig. 1).

Molecular data demonstrate convincingly the existence of the common lineage leading to the African apes and *Homo*. Along this common lineage (Fig. 1) numerous shared molecular changes have occurred. These involve derived amino acid sequence changes in the proteins albumin and transferrin, fibrinopeptides, carbonic anhydrase, haemoglobins, myoglobin, changes in mitochondrial DNA restriction sites, 1.5% base pair substitutions (involving $\sim 10^6$–10^7 base pair mutations) in nuclear DNA, and at least one locus with electrophoretic charge changes (refs 22, 23; Table 1). Specifically determined events and inferred molecular changes clearly document the existence of this lineage, and no molecular data reverse the position of *Pongo* and *Homo* shown in Fig. 1, although some data provide more reliable results than others. The conclusion, therefore, is that, in so far as the molecular evidence reveals phylogenetic relationships, the orang-utan is removed from unique association with the African apes and is the sister group to the clade comprising the African apes and man.

In addition to the branching pattern, the molecular data may also elucidate the timing of evolutionary events. It has been hypothesized that due to their apparent approximate regularity of change, molecules may serve as an evolutionary clock to date past divergence events[14,24], and this has been used to suggest that the human and African ape lineages separated as recently as 4–6 Myr ago. Calculations based on rates of nucleic acid and serum protein evolution[15] give dates ranging from ~8.5–14 Myr for the origin of the orang-utan lineage ($\bar{x} = 9.9 \pm 1.3$, $\gamma = 1.7$). An estimate of 10 ± 3 Myr is thus quite in order, given the approximate nature of the molecular clock[25], and this agrees with the fossil evidence described here for the origin of the orang-utan lineage. Previously, the fossil evidence has appeared to indicate dates of ~15–20 Myr for this event, but the data given here suggest that the date should be close to or slightly earlier than 10–11 Myr. Precluding naive interpretations of the molecular clock[25], these results are quite compatible with each other[26,27].

Morphological affinities among the hominoids

The evidence from gross morphology generally agrees with the molecular evidence, except that it defines the relationships among the extant hominoids less precisely. The evidence must be viewed at various levels. First, there is the evidence linking the hominoids as a monophyletic group through such shared derived characters as the presence of a vermiform appendix and the loss of the tail. Second, there are the numerous characters of the postcranial and axial skeleton which link the great apes and man and exclude the gibbons[28], particularly, for example, the morphology of the wrist[29], the structure and development of the shoulder musculature[30] and the reduced lumbar region of the vertebral column[31]. Finally, there are the characters that provide evidence of pairing relationships within the great apes and man, and as these are the principal concern of this paper these characters will be examined in more detail.

Evidence exists which previously led to the grouping of the great apes as a single group containing the chimpanzee, gorilla and orang-utan. These all have a diploid number of chromosomes of 48 compared with 46 in man. They have many superficial similarities in skull morphology, such as the prognathism of the face or the enlarged nuchal area of the occiput[32], and in their postcranial morphology, particularly in their limb proportions and shoulder morphology which were thought to be functionally adapted to brachiation[32]. More recently, most of these characters have been shown to be either allometric consequences of increased body size[28,33,34] or primitive characters retained by the living great apes[35] and therefore having no relevance to the within-group relationships of the living great apes.

There is much stronger evidence linking the African apes, the chimpanzees and gorillas with each other, to the exclusion of the orang-utan, including the many anatomical specializations related to knuckle walking in chimpanzees and gorillas[36], and other characters such as the presence of the frontal sinus[37,38], a character which they also share with man. This evidence is in general agreement with the molecular data, and there is little morphological evidence that lends any support to alternative sets of relationships, for example, that the orang-utan is related more closely to gibbons[39], that it forms the sister group to the group comprising the gibbons, African apes and man[39], or that the orang-utan by contrast is more closely related to man than are the African apes[40].

The lower face of chimpanzees and gorillas has many points of similarity with the human face despite the gross differences

Table 1 Molecular comparisons among the great apes and *Homo*

Molecular comparisons	*Homo–Pongo*	*Pan/Gorilla–Pongo*	*Homo–Pan–Gorilla*
Fibrinopeptide sequence (no. of amino acid differences)[23]	2	2	0
DNA hybridization (% base pair mismatch)[40]	2.4	2.0	1.1
DNA hybridization (% base pair mismatch)[23]	—	4.5	2.4
Immunological distance (ID units)[14]	37	39	14
Immunodiffusion distance (antigenic distance)[16,23]	1.6	1.6	0.8
Electrophoresis: plasma proteins[14,27]	»2	»2	~1.6
cladistic analysis based on number of electrophoretic charge changes at 23 loci[22]	8	9	7
genetic distance value based on 23 loci[17]	0.349	0.300	0.367
Mitochondrial DNA comparisons using endonuclease restriction enzyme maps of site shared[19]	1*	2†	7‡

* *Homo* shares a specific site with *Pongo* but others do not; † *Pan* and *Gorilla* share a site in common with *Pongo* but *Homo* does not; ‡ *Homo* shares this number of sites with at least one of the species of *Pan* or *Gorilla* but not with *Pongo* (or *Hylobates*).

Fig. 2 The maxilla of *S. meteai* showing the lower part of the face.

in facial height and the alveolar prognathism. Some of the characters shared between man, chimpanzees and gorillas are: the nasal aperture is broad; the subnasal plane is truncated and stepped down to the floor of the nasal cavity; the orbits are approximately square and often broader than high; the inter-orbital distance is broad; the infra-orbital foramina are usually three or less in number and are situated on or close to the zygomaticomaxillary suture; the zygomatic bone is usually curved and has a pronounced posterior slope; the zygomatic foramina (one or two) are small and are situated at or below the lower rim of the orbits; and the glabella is thickened (present in some fossil humans although not in modern man). The chimpanzee, gorilla and man also share the presence in the palate of small incisive foramina, large and oval-shaped greater palatine foramina, and large sphenopalatine fossae. Their teeth are basically similar in pattern except for the canines and premolars, and in particular the upper incisors lack a great size discrepancy between I^1 and I^2.

Many of the characters shared by man, chimpanzees and gorillas are present also in the gibbons. Where this is the case it is presumed that these characters are primitive for the Hominoidea, both because they are thus present in most extant hominoids and also because they are found in the most widely separated members of the group. (In other words, they are not present in the orang-utan, which separates the clades of gibbons and the African apes and man in Fig. 1.) In three of these characters, however, gibbons differ from the African apes and man, and these must be examined briefly to determine whether the primitive condition can be identified for them. Gibbons have a narrower oval-shaped nose, and from a consideration of other anthropoids, which generally have noses that are higher than broad, it seems likely that the primitive pattern for hominoids is a narrow oval-shaped nose as in gibbons. The infra-orbital foramina in gibbons are well removed from the zygomaticomaxillary suture, and again this is the usual condition in other anthropoids, thus it seems likely that this is the primitive condition in the hominoids. The condition of the glabella varies considerably in gibbons, from almost complete absence of thickening to fairly extensive thickening. This is true also of many other anthropoids, and it seems in fact to be size-related, therefore it is only possible to consider this character to be significant if glabellar thickening is either present on a small-bodied species or absent from a large one.

These characters have been summarized in a list (Table 2) which depicts the primitive condition in the Hominoidea. The two characters marked with an asterisk are present in gibbons, which thus retain the primitive condition, but not in the African apes and man, which thus share the derived conditions for these characters. All the other characters are shared between gibbons, African apes and man. One other character for which the polarity of evolutionary change is not yet understood concerns the presence of thick or thin enamel on the molars. The gibbons, chimpanzees and gorillas have thin enamel and man has thick enamel, and, on the face of it, thin enamel seems to be primitive for hominoids. Gibbons, chimpanzees and gorillas would thus have retained the primitive condition while man has developed the derived condition with thick enamel. Thick enamel, however, is a character that is shared with the orang-utan as well as with the fossil apes of the *Sivapithecus–Ramapithecus* group, and an understanding of the relationships between these two groups is essential to the understanding of human relationships.

Relationships of the orang-utan

The characters of the orang-utan differ in most respects from those listed in the previous section. Table 3 shows the characters of the orang-utan which it does and does not share with the fossil *S. meteai*[20] (Fig. 2). Other characters which it does not share with *S. meteai* could be added to the list, for example, the presence of wrinkling on the molar teeth in the orang-utan; but the absence of such clearly derived characters in the fossil hominoid need not be construed as evidence against an association between it and the orang-utan. Evidence for such an association comes from the list of shared characters, all of which (except where indicated) are interpreted as derived characters uniquely shared between *S. meteai* and the orang-utan. Where the fossil hominoid lacks the derived characters present in the orang-utan, this can be interpreted in terms of mosaic evolution, whereby some characters in an evolving lineage are developed before others. It would be more difficult to explain derived characters present in the fossil that are not present now in the orang-utan, but none have been observed.

No evidence exists as to whether the characters shared between *S. meteai* and the orang-utan are homologous or not. They may have arisen through convergent evolution rather than through common ancestry, but the latter is indicated by several sources. First, the characters shared between the orang-utan and *S. meteai* relate to separate functional complexes of the face and dentition. Some of the characters do inter-relate, such as the shape of the nose and the subnasal plane, or the nose and orbit shapes, or even possibly the shape of the nose and the narrow inter-orbital distance, but these characters relate directly neither to the characters of the palatine foramina, nor to two characters of the dentition, that is, the large difference in size between the central and lateral incisors and the thickness of enamel on the molar teeth. These combinations of characters can be seen as separate functional complexes that show similarities between the orang-utan and *S. meteai*, and if they are all the results of convergence it is difficult to envisage the circumstances that would lead to similarities in such disparate sets of characters occurring together in the fossil and extant forms.

Homology rather than convergence is also indicated by the metrical study of *S. meteai*[41], in which over a series of eight measurements of the palate and lower face, the extant hominoids showed a consistent and unified pattern differing

Table 2 Characters considered likely to be primitive for the Hominoidea

*Nasal aperture higher than broad, oval-shaped
Subnasal plane truncated
Subnasal plane stepped down to floor of nasal cavity
Orbits as broad as or broader than high
Inter-orbital distance broad.
Infra-orbital foramina few in number (≤3)
*Infra-orbital foramina well removed from the zygomaticomaxillary suture
Zygomatic bone curved and with strong posterior slope
Zygomatic foramina small
Zygomatic foramina 1–2 in number
Zygomatic foramina situated at or below the lower rim of the orbits
Glabella thickening which may occur on large individuals/species
Small incisive foramina
Large, oval-shaped greater palatine foramina
Upper incisors lacking large size discrepancy
Thin enamel on molars (see text)

This list is not exhaustive but concentrates on those characters that are known from the fossil record.
* These characters are present in gibbons but not in the African apes and man.

markedly from that of other anthropoids. The pattern for *S. meteai* approximates fairly closely to the great ape pattern, demonstrating both that its affinities in most dimensions are with the great apes and that, seen as a pattern, each of these characters is consistent with the others and therefore more likely to be the result of shared ancestry rather than independently acquired convergences. As for the non-metrical data, the variation in some of the dimensions, for example, palate length and nasal width, are not significantly inter-correlated and by inference are not functionally related.

Even more tenuously, it can be observed that in certain other respects the face of *S. meteai* bears strong resemblances to that of the orang-utan. The size and conformation of the zygomatic region, for example, are similar in the two hominoids, although the degree of similarity is not sufficient to distinguish them absolutely from the other hominoids. The great bizygomatic breadth in *S. meteai* is most similar to the condition in the orang-utan, but it also falls within the range of variation for male gorillas[20]. The same is true of the relatively great depth of the zygomatic region and the pronounced premaxillary prognathism[20]. The deep zygomatic region is combined in *S. meteai* with an estimated face length (nasal height measured from the estimated position of nasion[20]) that is below the range for the gorillas but within the range for the orang-utan. Thus these characters combine to produce a face shape similar to that of the orang-utan and different from comparable-sized gorillas. However, as the fossil is incomplete, it is not yet possible to quantify these characters adequately and little reliance can be placed on them except as support for the evidence given earlier.

It can be concluded from the morphological evidence given here that the orang-utan possesses a number of derived characters not present in the other extant hominoids. This is consistent with the molecular evidence discussed in the previous section (Fig. 1). We have shown that the suite of characters that distinguishes the orang-utan from other hominoids is present also in a fossil ape, *S. meteai*, and as it is concluded that these similarities are probably homologous we have felt justified in adding the fossil hominoid to the original cladogram to produce Fig. 3.

Sivapithecus indicus from India and Pakistan[3,11] can also be linked with *S. meteai* and the orang-utan on the basis of a newly described skull belonging to this species[42]. This skull, which was described after this review was written, shares with the orang-utan all the characters listed in Table 3, and it also provides evidence on two additional characters of the eye region which are included in Table 3. This strongly reinforces the case for the relationship of *Sivapithecus* with the orang-utan[44].

Discussion

S. meteai has long been identified as part of the *Sivapithecus–Ramapithecus* complex[3,20]. Simons and Pilbeam[3] synonymized

Table 3 Characters shared between the orang-utan and *S. meteai*

*Nasal aperture higher than broad, oval-shaped
Subnasal plane smooth, not truncated
Subnasal plane continuous with floor of nasal cavity, not stepped
Orbits higher than broad—not known for *S. meteai* but known for *S. indicus*[42]
Inter-orbital distance very narrow
*Infra-orbital foramina few in number
*Infra-orbital foramina well removed from the zygomaticomaxillary suture
Zygomatic bone flattened, facing anteriorly
Zygomatic foramina relatively large
No glabellar thickening—not known for *S. meteai* but known for *S. indicus*[42]
Extremely small incisive foramina
Slit-like greater palatine foramina
Great size discrepancy between I^1 and I^2
Thick enamel on molars
†Zygomatic foramen multiple (single in *S. meteai*)
†Zygomatic foramen situated above the level of the lower rim of the orbit (below the orbit in *S. meteai*)

* Characters considered to be shared primitive characters; all the rest are considered to be shared derived characters.

† Characters that are present in the orang-utan but for which *S. meteai* retains the primitive condition.

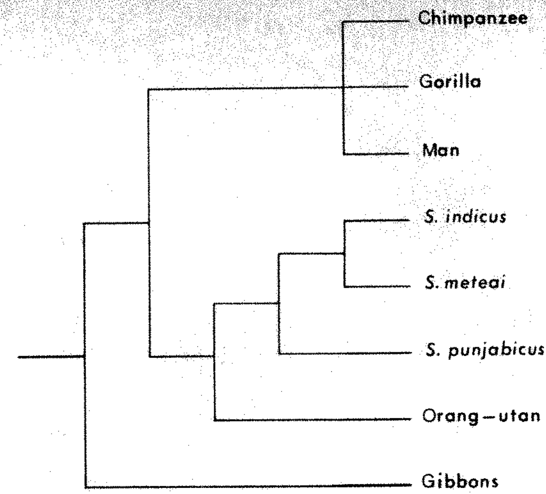


Fig. 3 Relationships of the extant and fossil Hominoidea. The relationships of the living species correspond to those in Fig. 1, and the fossils are shown as three species of *Sivapithecus* (including *Ramapithecus* as a junior synonym) which together constitute the sister group to the orang-utan.

it with *S. indicus*, but it was later reinstated as a separate species[20] on the basis of characters of the incisors, canines, premolars and posterior molars. In most of these characters it was presumed[20] that *S. meteai* possessed the derived condition with respect to the primitive hominoid morphotype and *S. indicus* the primitive condition. They both share, however, characters such as thickened enamel on the molar teeth, lower-crowned and more robust canines, and loss of molar cingula, which distinguish them from the primitive hominoid condition and indicate that the species presently assigned to *Sivapithecus* represent a clade, of which these are the distinguishing characters. As these characters are also shared by *Ramapithecus* it seems very probable that they are part of the same clade. This has been recognized implicitly by Pilbeam[11] in his combination of 'ramapithecines' and 'sivapithecines' in the family Ramapithecidae, and more explicitly by Greenfield[13] who dismissed even generic differences between them (although unfortunately he did so for the wrong reasons) and referred *Ramapithecus* to the genus *Sivapithecus*. This is certainly the simplest solution at present to a rather complicated situation, and although there is a risk that it is an over-simplification it does help to resolve many problems if the thick-enamelled Eurasian apes (excluding *Gigantopithecus*) are united in the genus *Sivapithecus*. This genus would then contain at least 3–4 species, which can be listed as follows with their geographical location and approximate ages:

S. indicus, India and Pakistan, middle to late Miocene, 11–8 Myr

S. punjabicus, India and Pakistan, middle to late Miocene, 11–8 Myr

Sivapithecus sp., China, late Miocene, 10–8 Myr

S. meteai, Turkey and Greece, late-middle Miocene, 11–10 Myr

Sivapithecus sivalensis[3] is of uncertain status, especially the type specimen of the species, and thus is omitted from the present discussion. The species of middle Miocene age assigned to *Sivapithecus* and *Ramapithecus*, *Sivapithecus darwini* from Turkey and Czechoslovakia[9] and *Sivapithecus africanus* and *Ramapithecus wickeri* from Kenya[8], are known only from extremely fragmentary dental remains and they lack the body parts that have been used in the present analysis to determine relationships. Therefore, these species also have been omitted from the present discussion.

Two consequences ensue from the recognition of the *Sivapithecus* group as a clade. First, if one member of the clade, that is, a species such as *S. meteai*, is accepted as being closely related to the orang-utan, then the other fossil species are also

linked in the same association. This would apply, for example, to species of *Ramapithecus* if they are congeneric with *Sivapithecus*. Only if characters are found to be present in more complete specimens that exclude them from this association (and also of course from the genus *Sivapithecus*) can it be accepted that they do not form part of this clade. Following this line of argument, in conjunction with the molecular and morphological evidence for the relationship of the extant apes, it is evident that '*Ramapithecus*' may not be related exclusively to man but may be related to the orang-utan.

The case for the hominid affinities of ramapithecines has most recently been reviewed by Kay[43]. He also refers '*Ramapithecus*' to *Sivapithecus*, although not as a separate species but to the species *S. sivalensis*. He shows that the species of *Sivapithecus* share the following synapomorphies with man: (1) broad mandibular body; (2) low-crowned molars with thick enamel; (3) rather reduced canine size; (4) rather reduced canine sexual dimorphism; (5) buccolingually broad upper canines; and (6) tendency to enlarged P_3 metaconids. These are said to be characteristic of the whole *Sivapithecus* clade and that this is therefore the sister clade of man. We accept Kay's interpretation of the significance of characters 1 and 5, although mandibular robusticity varies considerably for the different species of *Sivapithecus*, but there are greater difficulties with the others. Two of the characters, the enlarged P_3 metaconids and the thick enamel of the molars, are present also in the orang-utan, and therefore provide equally good evidence for a *Sivapithecus–Pongo* relationship as for a *Sivapithecus–Homo* relationship. We do not accept Kay's conclusion that canine size is reduced in *Sivapithecus* for two reasons; first, his data show that the range of canine size relative to the first molar is equivalent to the middle of the range of extant great apes: in other words *Sivapithecus* canines are relatively bigger than female great ape canines as well as being relatively smaller than male canines. Secondly, it must be considered that *Sivapithecus* may have relatively large molars for its body size[11], and if this is the case the ratio of canine/molar size may be misleading. Finally, we also disagree with the significance attached to reduced sexual dimorphism, as low sexual dimorphism has not been shown convincingly to be a valid characteristic of *Sivapithecus*.

The characters linking *Sivapithecus* with man are thus reduced to two: robust mandibles and buccolingually broad upper canines, two characters that are almost certainly functionally related. These characters certainly seem to be synapomorphies for *Homo* and *Sivapithecus*, but the likelihood of their being the result of parallel evolution must be given credence in view of the more numerous and functionally independent derived characters of the nose, the orbits, the palatal foramina and the incisors that are shown here to be shared between *Sivapithecus* and *Pongo*. We therefore consider it more likely that *Sivapithecus* species, particularly *S. meteai* and *S. indicus* but by inference also *Sivapithecus* (formerly *Ramapithecus*) *punjabicus*, are more closely related to the orang-utan than to man.

The other consequence arising from the proposed relationship of *S. meteai* and the orang-utan is that it provides an approximate minimum divergence data between the orang-utan and its extant sister group, the African apes and man. *S. meteai* itself comes from deposits about 10–11 Myr old, but the time span of *Sivapithecus* as recognized here is of the order 8–11 Myr. This age corresponds to the date of 10 ± 3 Myr suggested by the molecular evidence and strongly indicates a minimum age of at least 10–11 Myr for the divergence between the orang-utan clade and the clade containing the African apes and man. The divergence between the members of the latter must be subsequent to this period, and the molecular evidence for a prolonged common ancestry between the African apes and man suggests that their eventual divergence was considerably later than the middle Miocene and perhaps as late as the end of the Miocene 5–6 Myr ago[15].

There is one other consequence of the relationships proposed here, which will be considered briefly. The classification of the cladistic groups proposed in Fig. 3 would differ greatly from many current schemes. The most recent one by Szalay and Delson[5], for example, combines the great apes in the tribe Pongini with parallel tribes for *Dryopithecus* (Dryopithecini) and *Sivapithecus* (Sugrivapithecini). These are combined in the subfamily Ponginae, and *Ramapithecus*, *Australopithecus* and man are classified in the subfamily Homininae. These subfamilies are then grouped in the family Hominidae. Classifications based on Fig. 3 would have to recognize the chimpanzee, gorilla and man as one clade with the orang-utan and *Sivapithecus* (and '*Ramapithecus*') as another; a possible solution would be to classify these clades at subfamily level, as do Szalay and Delson[5], but with Homininae for the African apes and man, and Ponginae for orang-utan and *Sivapithecus* (and '*Ramapithecus*'). The African apes and man would then be divided at the level of tribe. An alternative classification at the family level is as follows:

- Hominoidea (Gray, 1825)
  - Hylobatidae (Blyth, 1875)
    - *Hylobates* (Illiger, 1811)
  - Pongidae (Elliot, 1913)
    - *Pongo* (Lacepede, 1799)
    - *Sivapithecus* (Pilgrim, 1910)
  - Hominidae (Gray, 1825)
    - Gorillinae (Hurzeler, 1968)
      - *Pan* (Oken, 1816)
      - *Gorilla* (I. Geoffroy, 1852)
    - Homininae (Gray, 1825)
      - *Homo* (Linnaeus, 1758)
      - *Australopithecus* (Dart, 1925)

This change from traditional classifications has long been advocated by Goodman[23], and it is one that is necessary with the recognition that the African apes are more closely related to man than they are to the orang-utan.

We thank N. Boaz, L. Brunker, E. Delson, S. Gould, T. Harrison, J. Lowenstein, L. Martin, E. Meikle and C. Stringer for helpful comments. V. Sarich provided helpful discussion on the molecular section of this paper, and much of the morphological evidence is the product of an ongoing collaboration with C.P. Groves. The Wenner-Gren Foundation, the L.S.B. Leakey Foundation and the Milton Fund of Harvard University provided grants to J.E.C. in support of part of this work.

1. Pilbeam, D. R. *A. Rev. Anthrop.* **8**, 333–352 (1979).
2. Andrews, P. in *Aspects of Human Evolution*, 25–61 (Taylor & Francis, 1981).
3. Simons, E. L. & Pilbeam, D. R. *Folia Primatol.* **3**, 81–153 (1965).
4. Delson, E. in *Paleoanthropology in The Peoples Republic of China* 40–65 (Natn. Acad. Sci., Washington D.C., 1977).
5. Szalay, F. S. & Delson, E. *Evolutionary History of the Primates* (Academic, New York, 1979).
6. Howell, F. C. in *Evolution of African Mammals*, 154–248 (Harvard University Press, 1978).
7. Johanson, D. C. & White, T. D. *Science* **203**, 321–330 (1979).
8. Andrews, P. & Walker, A. *Human Origins*, 279–304 (Benjamin, Menlo Park, 1976).
9. Andrews, P. & Tobien, H. *Nature* **268**, 699–701 (1977).
10. Xu, Q. & Lu, Q. *Vertebr. palaeont.* **17**, 1–13 (1979).
11. Pilbeam, D. R. *et al. Nature* **270**, 689–695 (1977).
12. Simons, E. L. *Phil. Trans. R. Soc.* **292**, 21–41 (1981).
13. Greenfield, L. O. *Am. J. phys. Anthrop.* **50**, 527–546 (1979).
14. Sarich, V. M. & Cronin, J. E. in *Molecular Anthropology*, 141–170 (Plenum, New York, 1976).
15. Cronin, J. E. & Meikle, W. E. *Int. J. Primatol.* (in the press).
16. Goodman, M. in *Molecular Anthropology*, 321–353 (Plenum, New York, 1976).
17. Bruce, E. & Ayala, F. *Evolution* **33**, 1040–1056 (1979).
18. Smith, R. J. & Pilbeam, D. R. *Nature* **284**, 447–448 (1980).
19. Ferris, S. O., Wilson, A. C. & Brown, W. M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2432–2436 (1981).
20. Andrews, P. & Tekkaya, I. *Palaeontology* **23**, 85–95 (1980).
21. Lowenstein, J., Sarich, V. M. & Richardson, B. J. *Nature* **291**, 409–411 (1981).
22. Cronin, J. E. in *New Interpretations of Ape and Human Ancestry* (Plenum, New York, in the press).
23. Goodman, M. & Cronin, J. E. in *A History of Physical Anthropology 1930–1980* (Academic, New York, in the press).
24. Wilson, A. C., Carlson, S. S. & White, T. J. *A. Rev. Biochem.* **46**, 573–630 (1977).
25. Corruccini, R., Baba, M., Goodman, M., Ciochon, R. L. & Cronin, J. E. *Evolution* **34**, 1216–1219 (1980).
26. Zihlman, A. L., Cronin, J. E., Cramer, D. L. & Sarich, V. M. *Nature* **275**, 744–746 (1978).
27. Cronin, J. E. & Meikle, W. E. *Syst. Zool.* **28**, 259–269 (1979).
28. Andrews, P. & Groves, C. P. in *Gibbons and Siamang* Vol. 4, 167–218 (Karger, Basel, 1975).
29. Lewis, O. J. *Am. J. phys. Anthrop.* **30**, 251–268 (1969).

30. Ashton, E. H. & Oxnard, C. E. *Trans. zool. Soc. Lond.* **29**, 553–650 (1963).
31. Schultz, A. H. *Primatologia* **4**, 5/1–5/66 (1961).
32. Le Gros Clark, W. E. *The Fossil Evidence for Human Evolution* (University of Chicago Press, 1955).
33. Aiello, L. in *Aspects of Human Evolution*, 63–97 (Taylor & Francis, 1981).
34. Biegert, J. & Maurer, R. *Folia Primatol.* **17**, 142–156 (1972).
35. Delson, E. & Andrews, P. in *Phylogeny of the Primates*, 405–446 (Plenum, New York, 1975).
36. Tuttle, R. H. *Am. J. phys. Anthrop.* **26**, 171–206 (1967).
37. Cave, A. J. E. & Haines, R. W. *J. Anat.* **74**, 493–523 (1940).
38. Cave, A. J. E. *Proc. zool. Soc. Lond.* **136**, 359–373 (1961).
39. Romero-Herrara, A. E., Lehmann, H., Castillo, O., Joysey, K. A. & Friday, A. E. *Nature* **261**, 162–164 (1976).
40. Benveniste, R. E. & Todaro, G. J. *Nature* **261**, 101–108 (1976).
41. McHenry, H. M., Andrews, P. & Corruccini, R. S. *Folia Primatol.* **33**, 241–252 (1980).
42. Pilbeam, D. R. *Nature* **295**, 232–234 (1982).
43. Kay, R. F. *Int. J. Primatol* (in the press).
44. Andrews, P. *Nature* **295**, 185–186 (1982).

ARTICLES

Ancient and modern slopes in the Tharsis region of Mars

P. J. Mouginis-Mark*, S. H. Zisk† & G. S. Downs‡

* Department of Geological Sciences, Brown University, Providence, Rhode Island 02912, USA
† NEROC Haystack Observatory, Westford, Massachusetts 01886, USA
‡ Jet Propulsion Laboratory, Pasadena, California 91109, USA

The directions of lava flows in the Tharsis region of Mars are used to identify regional palaeo-slopes, vent areas and local topography. A comparison is made between these flow directions and the present day radar-measured topography; good agreement between these data sets indicates that little intra-regional tectonic deformation has occurred following the emplacement of the preserved lavas.

AS the youngest centre of constructional volcanism on Mars[1,2], the Tharsis region has been the focus of considerable attention directed towards gaining an understanding of the geophysical[3,4], geochemical[5] and volcanological[6] evolution of the area. Debate has centred on the relative importance of regional tectonic uplift versus constructional volcanism as the primary mechanism for the production of the observed relief[7,8], together with the attendant implications that such hypotheses would hold for lithospheric structure and crustal–mantle petrology. Until recently, it was believed that the centre of the broad Tharsis dome was located close to the three Tharsis Ridge volcanoes, where the surface rose to nearly 10 km above the martian datum[9,10]. New Earth-based radar measurements have shown, however, that Tharsis is much lower on average, with the mean elevation only 2–4 km above the 6.1 mbar reference surface[11]. In addition, the centre of the dome is evidently located further to the south-east than previously believed and is now centred on Syria Planum[12].

Nevertheless, considerable local topography (including numerous volcanic constructs) exist within Tharsis by virtue of the many outpourings of lava which characterize the area. As a consequence of the changing lithospheric load associated with this volcanic activity, deformation of areas such as the base of Olympus Mons would be expected if the lithospheric thickness was $\leq$150 km (ref. 13). In an attempt to search for recent tectonic deformation of this kind within Tharsis, we compare here the ancient slope directions (deduced from the orientations of lava flows) with the present-day slopes (derived from Earth-based radar measurements). Due to their propensity to flow parallel to the direction of slope, the eruptions of lava provide information on the characteristics of the local topography at the time of flow emplacement. In addition, the travel directions and areal extent of the flows can be used to describe further the nature of the volcanic activity within Tharsis because they permit the different eruptive centres and their relative volumes of lava to be recognized.

Flow measurements

Individual lava flows within Tharsis were mapped either from high resolution (20–150 m per pixel) Viking Orbiter images or from the new medium resolution (200 m) orthophotomosaics prepared by the US Geological Survey. The positions of these lava flows are shown in Fig. 1, with the exception of those flows which are located on the high flanks of the shield volcanoes. A few flow units in addition to those mapped may also exist, particularly to the west of Ulysses Patera and north of Noctis Labyrinthus, where only poor photographic coverage exists due to atmospheric opacity; some lobate flow edges are visible in these regions, but the boundaries cannot be traced for sufficient distances to infer their direction of travel.

A total of 475 lava flows longer than 20 km have been mapped (Fig. 1). These flows have typical widths of 5–10 km on the higher slopes of the volcanoes and 15–35 km on the lower flanks. Additional lava flows that spread out over wide areas and have widths greatly in excess of 40 km were deliberately excluded from this analysis because they were deemed unsuitable for the determination of their flow directions. In certain instances, particularly to the west of Arsia Mons, it has also proved difficult to map the total extent of each flow, since many flow segments are either partially buried by superposed younger units or their boundaries have become indistinct as more subdued relief is encountered close to their vents. Thus, while the flows measured here rarely exceed 200 km in length, previous estimates of 300–400 km appear to be quite plausible[14,15], while flow length maxima of ~1,000 km may also be possible for fissure eruptions originating at low elevations. For the flows included within this analysis, independent estimates of the flow thicknesses made from shadow measurements[14] indicate values between 5–20 m for the steeper slopes and 20–65 m for the generally flatter surrounding terrain.

The map of Tharsis lava flows (Fig. 1) enables both the regional slopes and the individual eruptive centres at the time of flow emplacement to be recognized. In agreement with the general observations of Schaber *et al.*[14], our analysis shows that the major palaeo-slope within Tharsis was down towards the north-west, with a high point present between Arsia and Pavonis Montes and a low located near to Olympus Mons. From the distribution of the flows, the linear extent of this north-west slope was at least 2,500 km, and there were also downhill gradients which extended away from Pavonis and Ascraeus Montes towards the north-east for 2,000 km, south-east from Syria Planum for 1,200 km and westward from Arsia Mons for ~1,000 km. While most of the flows originated from the four major shield volcanoes[14–16], additional lava flows were

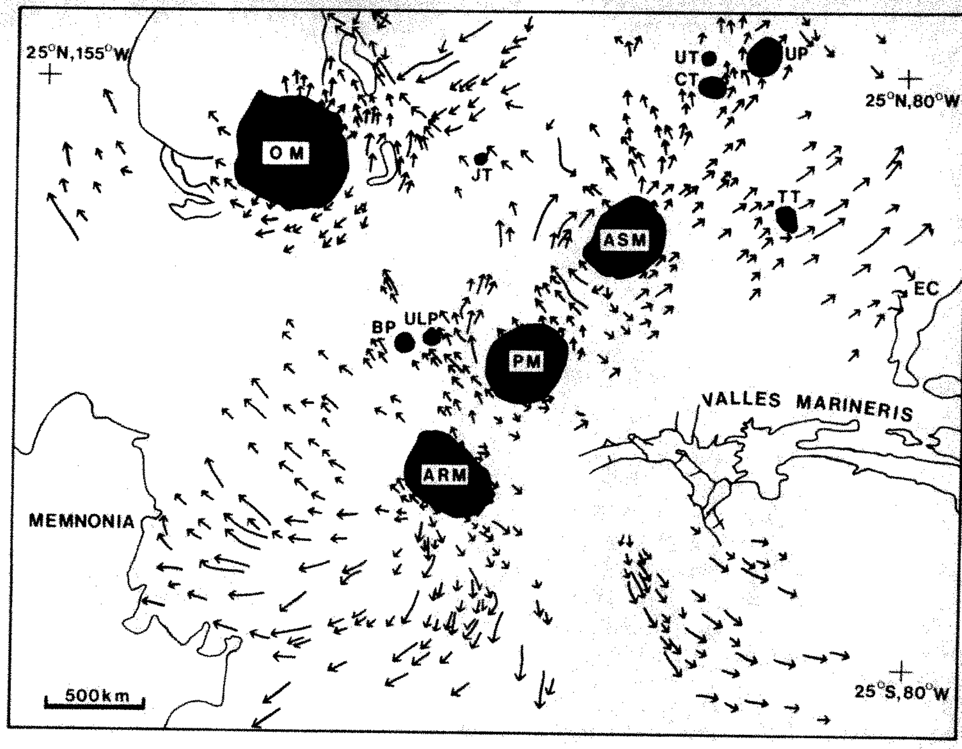


Fig. 1 Distribution of all lava flows included within this analysis. Length of arrow is equal to the length of the lava flow and shows the direction of travel. Volcanoes are shown in black where: ARM, Arsia Mons; ASM, Ascraeus Mons; BP, Biblis Patera; CT, Ceraunius Tholus; JT, Jovis Tholus; OM, Olympus Mons; TT, Tharsis Tholus; ULP, Ulysses Patera; UP, Uranius Patera; UT, Uranius Tholus. Also shown are the highland boundary in Memnonia (lower left), Valles Marineris, the Olympus Mons aureole (north-west of the volcano) and Echus Chasma (EC).

erupted in areas which lack obvious volcanic constructs (Fig. 1); Syria Planum and an area to the west of Ceraunius Fossae (25°N, 110°W) are two such examples. In these localities, several small cones of probable volcanic origin exist[17], but their significance as source regions for large volumes of magma had not previously been recorded. Lava flows >100 km in length originate from each of these two areas, indicating the existence of well developed fissure systems which were distinct from the conduits which supplied the magma to the major volcanoes.

Local deformations

In addition to illustrating the regional slopes within Tharsis, the orientations of the lava flows permit deformational features on a scale of a few hundred kilometres to be identified. Prominent amongst these is a circumferential depression, or peripheral trough, which surrounds the southern half of Olympus Mons[11]. Extending to distances of 300–350 km beyond the basal escarpment of the volcano, this depression has acted as a funnel to redirect lavas which consequently either flowed northward towards Arcadia Planitia or westward towards Amazonis Planitia. Segments of the aureole material to the east of Olympus Mons have been embayed and partially buried by lavas from both central Tharsis and Olympus Mons, while the much more extensive portions of the aureole to the north and west of the volcano, despite being down the regional slope[12], have remained unburied. From our investigation of the preserved record of volcanic activity, none of the lava flows erupted from Olympus Mons contributed to the lava pile which now constitutes central Tharsis (such as the area around Biblis and Ulysses Paterae); all the lava flows from the volcano travelled either to the north or west. This implies that the peripheral trough around Olympus Mons has been in existence for an extended period of time, rather than being a relatively recent response to crustal loading by the volcano after most of the lavas had been emplaced.

In contrast to Olympus Mons, the other three Tharsis shield volcanoes are noteworthy for the lack of preserved deformational features within the surrounding plains materials. A slight topographic rise ~330 km west of Ascraeus Mons, together with the tangential flow direction of lavas at the base of the volcano, suggest that a local depression may exist around this volcano, although subsequent infilling by lavas makes this identification uncertain. No similar evidence exists from the orientations of preserved lava flows for depressions around either Arsia or Pavonis Montes; flows from these two volcanoes travelled in almost radial directions from the summit calderas and adjacent fissure zones.

Only one example of surface deformation after lava flow emplacement can be found in the Tharsis region. To the north-east of the Olympus Mons peripheral trough, relatively old lava flows from the south-east have been partially buried and modified by more recent volcanism. In one instance, a sinuous lava channel was formed along a radial directed towards Olympus Mons and this lava channel cuts the older flows almost at right angles to the original flow direction (Fig. 2). Our interpretation of this dual flow direction is that after their emplacement, the older flows were tilted towards the south-west before the formation of the lava channel. Lithospheric loading of the area by the continuing construction of Olympus Mons could be the cause of this tectonic movement.

Present-day topography

Earth-based radar measurements of elevations on Mars have recently been referenced to the 6.1-mbar pressure surface[11].

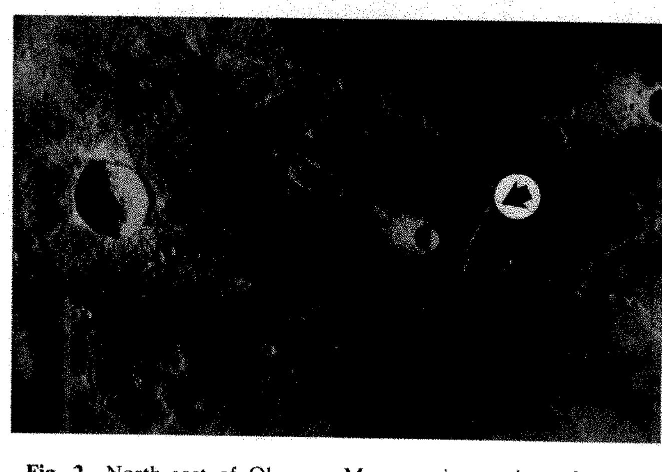


Fig. 2 North-east of Olympus Mons, a sinuous lava channel (source arrowed) cuts older lava flows which came from the south-east. The change in flow direction between these two eruptions is attributed to lithospheric loading associated with the construction of Olympus Mons. North is towards the top of image, which is centred at 29.8° N, 130.0° W. Image width is 60 km; Viking Orbiter frame 623A46.

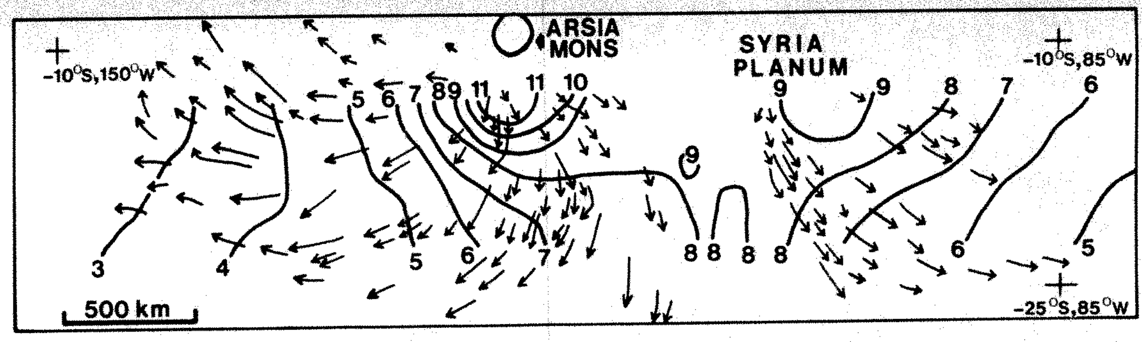


Fig. 3 Comparison of radar-derived topography and lava flow directions for the southern portion of the study area. Radar data give contours in 1-km intervals and are referenced to the 6.1-mbar pressure surface of Mars. Contours were hand drawn (by P.M.M.) from elevation measurements made in 1971 and 1973 between 14° and 21° S.

These data provide height information for the southern flanks of Arsia Mons and Syria Planum for latitudes 14–21°S. Data resolution for this area is ~10 km in longitude, 80 km in latitude and 100 m in altitude[18]. Using this radar-derived topography, it is therefore possible to compare the present day regional gradients with the observed flow directions (palaeo-slopes) to search for regional deformations which may have occurred after the cessation of the volcanic activity. Such tectonic deformation would be expected to be present if the lithosphere beneath Tharsis was thinner than ~150–180 km (refs 13, 19) due to the changing load associated with the protracted eruption of large volumes of lava[20,21].

The following comparison of these present and past slopes within southern Tharsis (Fig. 3) points to a strong agreement between the two data sets. The radar elevations corroborate that a regional slope extends downward from Arsia Mons towards Memnonia; the total change in elevation amounts today to >8 km. A slight (500–1,000 m) high is associated with the Claritas Fossae fractures, while Syria Planum represents the centre of a regional dome located at the western end of Valles Marineris[12]. Almost all the mapped lava flows cross the present day contour lines within about 10° of the local gradient (Fig. 3). Considering that the aforementioned resolution of the radar data provides only the general characteristics of the local topography, there is also good agreement between the regional slopes and the paths of the flows (such as those associated with the 'bending' of the flows south of Arsia Mons between latitudes 17°–22°, longitudes 110°–122°W). Essentially all of the observed lava flows are seen to travel generally down the present topographic gradient for their entire length. Close to the summit of Arsia Mons, however, where a parasitic shield is thought to be located[22], some flows do not follow the maximum slope. This may, on the other hand, be a consequence of the shallowness of the slopes (see below) and the relative thickness of the flows (~25 m (ref. 14)), which could have funnelled the younger flows around pre-existing units in a manner comparable to the diversion of certain terrestrial eruptions (ref. 23, p. 421).

Although the radar data set lacks sufficient areal coverage to compile a complete topographic map, scattered elevation values for parts of northern Tharsis are available[11] and permit the determination of the present-day slopes north of the equator. We have measured 14 slopes within Tharsis (all of which parallel the lava flow directions for distances of 150–1,250 km) using the radar data (Fig. 4). In each case, it is evident from the slope estimates (Table 1) that the lava flows travelled along paths that are consistent with the modern down-slope directions. All the radar-measured slopes seem to be very shallow, ranging from 0.75° for the southern flanks of Arsia Mons (profile L–L′) to 0.04° for a section of the Olympus Mons peripheral trough (profile C–C′). These slopes are, however, only best estimates from the radar data because each elevation estimate is an average of several measurements made within a 2° longitude sample bin[11]. Typically, 7–13 measurements are included in each sample, and the resulting standard errors for the averages range from 70 to 770 m (Table 1). Over the measured horizontal distances considered here, the maximum error in the slope determination would be about ±0.17°, while typical errors would be <±0.1°. Under these circumstances, only in the case of profile C–C′ could the down-slope direction of the profile change; although the absolute slopes would be different, all the remaining modern gradients are consistent with a down-slope movement of the lava flows even allowing for the resolution of the radar system.

Such a close match between the ancient and modern regional slopes indicates appreciable lithospheric stability within Tharsis. Although the absolute ages of the lava flows are not known, at least 14 distinct eruptive sequences of flows have been identified in this area[14] and these surface units have large differences in their superposed crater densities[21]. Crater counts give number densities ranging from 1.22×10^{-3} to 9.08×10^{-5} craters >1 km diameter per km^2 respectively for the Tharsis Montes unit 3 (tm_3) and Olympus Plains (op) units of Scott and Tanaka[21], suggesting a wide age range. It appears, therefore, that there was relatively little tilting of geological units during (or after) the extended period of Tharsis volcanism represented by the currently exposed lava flows. This in turn implies that the lithospheric structure beneath Tharsis was rigid enough to support the changing load associated with volcano construction, suggesting that the lithosphere may indeed have been thicker than ~150–180 km (refs 13, 19, 20).

Volcanology

In addition to the slope information, subtle differences in the style of Tharsis volcanism are also indicated by the distribution of the observed lava flows. From the large number of flows located between the three Tharsis Ridge volcanoes (Fig. 1), it

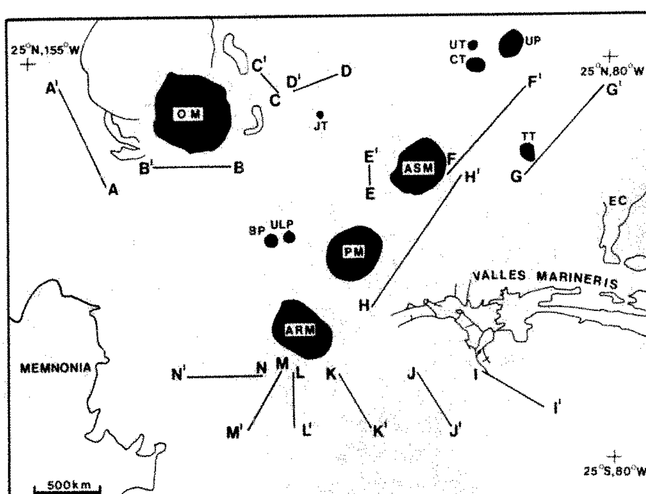


Fig. 4 Location of the surface slopes measured from Earth-based radar data (see Table 1). Each slope is coincident with the lava flow directions shown in Fig. 1, with the downslope measurement denoted by the superscript bar. Base map abbreviations are the Hsame as those in Fig. 1.

Table 1 Radar-derived measurements of modern slopes in Tharsis region

	Up-slope (unprimed) Lat. (deg)	Long. (deg)	Altitude* (km)	S	N	Down-slope (primed) Lat. (deg)	Long. (deg)	Altitude* (km)	S	N	Distance† (km)	Height‡ (km)	Slope§ (deg)
A–A′	9.41	144–146	−0.49	0.15	9	22.00	150–152	−2.02	0.09	12	837	1.53	0.10±0.02
B–B′	12.11	128–130	2.22	0.10	11	12.11	138–140	0.07	0.18	9	600	2.15	0.20±0.03
C–C′	21.58	122–124	2.35	0.07	12	23.15	124–126	2.25	0.09	12	152	0.10	0.04±0.07
D–D′	23.15	114–116	4.45	0.14	12	21.58	120–122	2.65	0.13	13	372	1.80	0.28±0.04
E–E′	9.61	110–112	4.43	0.23	11	12.11	110–112	3.65	0.22	7	150	0.78	0.30±0.17
F–F′	10.85	100–102	4.90	0.25	3	21.78	90–92	3.47	0.08	9	889	1.43	0.09±0.02
G–G′	10.98	90–92	3.55	0.15	8	21.78	80–82	1.76	0.10	13	883	1.79	0.12±0.02
H–H′	−6.10	110–112	8.62	0.09	2	10.98	98–100	4.64	0.13	5	1252	3.98	0.18±0.01
I–I′	−14.47	96–98	8.74	0.10	13	−18.99	88–90	6.04	0.10	12	551	2.70	0.28±0.02
J–J′	−14.66	104–106	8.29	0.30	13	−21.11	100–102	7.63	0.10	12	455	0.66	0.09±0.05
K–K′	−14.47	114–116	8.74	0.05	12	−21.11	110–112	7.44	0.08	13	455	1.30	0.16±0.02
L–L′	−14.47	120–122	11.75	0.08	13	−21.11	120–122	6.69	0.12	12	386	5.06	0.75±0.03
M–M′	−14.47	122–124	10.98	0.57	12	−21.23	126–128	5.07	0.12	13	471	5.91	0.72±0.08
N–N′	−14.47	124–126	8.29	0.77	13	−14.46	134–136	4.32	0.13	12	600	3.97	0.38±0.09

* Altitudes are expressed in kilometres above the 6.1-mbar pressure surface. 1 σ (S) values from the mean elevation and the number of data points (N) in the two-degree sample bin are also given.

† Straight-line distances are given from the middle of the up- and down-slope points.

‡ Height difference between the mean up- and down-slope altitudes.

§ Surface slopes, estimated errors for these slope values are derived from the 1σ values in the altitude measurements.

is evident that large volumes of magma were also erupted at these localities from saddle fissures (in addition to the magma erupted from the summit calderas). Furthermore, a change in eruptive style from true shield volcanism (most of the magma erupted from near-summit vents) to flank activity (most of the magma finding egress at the saddle crests) evidently took place on both Ascraeus and Pavonis Montes late in their evolution, because no flows from their summit calderas extend across the saddle flows. Eruptions from Arsia Mons also display a similar change in vent location, with large eruptions of magma from the parasitic cone[22] postdating the near-summit activity. This sequence of vent relocations for these volcanoes is at least qualitatively consistent with the inferred history of the calderas of the Tharsis volcanoes[24].

All the small volcanic constructs of the Tharsis region (Biblis, Ulysses and Uranius Paterae and Tharsis, Uranius, Jovis and Ceraunius Tholi) show evidence of partial burial by lava flows which conform to the regional slopes identified in this analysis[25]. The absence of minor deviations from a straightline path, which most of the flows follow, indicates that there are no additional topographic rises within central Tharsis which could be other low volcanoes. Thus, it seems that the entire sequence of lava plains in this region[21] was constructed by a few major volcanic centres (not all of which were associated with the large shields); eruptions of magma were evidently confined to large, frequently reactivated fissure systems, rather than many small widely dispersed vents, during the period of activity presently exposed at the surface.

Comparing the size and number of the flows originating from Arsia Mons (42 > 100 km long) and Olympus Mons (3 > 100 km long) indicates that it was likely that these two volcanoes had appreciably different magma supply rates from their respective source regions. If it is assumed that these martian shields had eruption characteristics comparable to terrestrial volcanoes (that is, lava flow length was related to magma effusion rate[26]), the longer and more numerous flows from Arsia Mons would imply higher effusion rates than for Olympus Mons. Variations in effusion rate are also likely for different parts of Olympus Mons, because relatively long flows are only common on the eastern and southern flanks of the volcano while the aureole material to the north-west is almost devoid of embaying lavas. Thus, the traditional view that Olympus Mons is the largest volcano in the Solar System[1] may not be correct, because Arsia Mons not only has a larger caldera, but from this analysis also appears to have higher magma effusion rates and represents a topographic high that dominates the surrounding region for radial distances >1,000 km in any direction from its caldera. In comparison to both Arsia and Olympus Montes, the mass eruption rates for individual flows from Pavonis (seven flows longer than 100 km) and Ascraeus (nine flows longer than 100 km) were probably intermediate between these two extremes.

Conclusions

The use of lava flow directions as palaeo-slope indicators and Earth-based radar measurements of present topography demonstrate that little deformation of the Tharsis region has occurred over an extended period of time. Very few deformational features attributable to subsidence around the volcanoes can be recognized from the flow directions; a peripheral trough surrounds Olympus Mons and there is some evidence for a similar feature to the west of Ascraeus Mons. Only one example of surface tilting after lava emplacement has been found (to the north-east of Olympus Mons), suggesting that the increasing lithospheric load due to progressive buildup of the lava pile had little influence on local topography. It is therefore postulated that early in the evolution of Olympus Mons, the lithosphere was <150 km thick[13], thereby permitting the formation of the peripheral trough. For much of the period in which the preserved volcanic activity took place, however, lithospheric thickness has evidently been >150–180 km (refs 13, 19), as evidenced by the lack of local or regional deformation between the time of these eruptions and the present day.

Although regional slopes are very shallow today, if they truly reflect the gradients present at the time of volcanic activity, they nevertheless controlled lava flow directions over distances in excess of 1,000 km. In this respect, the rise associated with Syria Planum[12] has controlled the flow directions of many of the Tharsis lavas, while Claritas Fossae was also a local high at the time that Arsia Mons was active. It is evident that the style of volcanic activity was quite diverse within the region, with fissure eruptions from the saddles between the Tharsis Ridge volcanoes contributing large volumes of lava to the surrounding plains. In this connection, it is proposed that Arsia Mons, with its extensive lava flows and great relief, should supersede Olympus Mons for the title of the largest known volcano in the Solar System.

We thank Jim Head for his comments. This research was supported by NASA grant NER 40-002-088 of the Planetary Geology Program.

Received 18 February; accepted 1 April 1982.

1. Mutch, T. A., Arvidson, R. E., Head, J. W., Jones, K. L. & Saunders, R. S. *The Geology of Mars* (Princeton University Press, 1976).
2. Scott, D. H. & Carr, M. H. *Geologic Map of Mars* (U.S. Geological Survey Miscellaneous Map I-1083, 1978).
3. Sleep, N. H. & Phillips, R. J. *Geophys. Res. Lett.* **6,** 803–806 (1979).
4. Phillips, R. J. & Lambeck, K. *Rev. Geophys. Space Phys.* **18,** 27–76 (1980).
5. Finnerty, A. A. & Phillips, R. J. *Abstr. 3rd int. Colloq. Mars* 77–79 (Contr. 441, Lunar and Planetary Institute, Houston 1981).
6. Plescia, J. B. & Saunders, R. S. *Proc. 10th Lunar planet. Sci. Conf.* 2841–2859 (1979).

7. Plescia, J. B. & Saunders, R. S. *Proc. 11th Lunar planet. Sci. Conf.* 2423–2436 (1980).
8. Solomon, S. C. & Head, J. W. *Lunar planet. Sci.* **11**, 1063–1065 (1980).
9. Christensen, E. J. *J. geophys. Res.* **80**, 2909–2913 (1975).
10. *Topographic Map of Mars* (U.S. Geological Survey Miscellaneous Map I-961, 1976).
11. Downs, G. S., Mouginis-Mark, P. J., Zisk, S. H. & Thompson, T. W. *J. geophys. Res.* (in the press).
12. Downs, G. S., Mouginis-Mark, P. J., Zisk, S. H. & Thompson, T. W. *Lunar planet. Sci.* **13**, 182–183 (1982).
13. Thurber, C. H. & Toksöz, M. N. *Geophys. Res. Lett.* **5**, 977–980 (1978).
14. Schaber, G. G., Horstman, K. C. & Dial, A. L. *Proc. 9th Lunar planet. Sci. Conf.* 3433–3458 (1978).
15. Carr, M. H., Greeley, R., Blasius, K. R., Guest, J. E. & Murray, J. B. *J. geophys. Res.* **82**, 3985–4015 (1977).
16. Crumpler, L. S. & Aubele, J. C. *Icarus* **34**, 496–511 (1978).
17. Moore, H. J. & Hodges, C. A. *NASA-TM 82385* 266–268 (1980).
18. Downs, G. S., Reichley, P. E. & Green, R. R. *Icarus* **26**, 276–312 (1975).
19. Turcotte, D. L., Willemann, R. J., Haxby, W. F. & Norberry, J. *J. geophys. Res.* **86**, 3951–3959 (1981).
20. Comer, R. P., Solomon, S. C. & Head, J. W. *Lunar planet. Sci.* **11**, 171–173 (1980).
21. Scott, D. H. & Tanaka, K. L. *Icarus* **45**, 304–319 (1981).
22. Roth, L. E., Downs, G. S., Saunders, R. S. & Schubert, G. *Icarus* **42**, 287–316 (1980).
23. Macdonald, G. A. *Volcanoes* (Prentice-Hall, New Jersey, 1972).
24. Mouginis-Mark, P. J. *Proc. 12th Lunar planet. Sci. Conf.* 1431–1447 (1981).
25. Pike, R. J. & Clow, G. D. *Abstr. 3rd int. Colloq. Mars* 199–201 (Contr. 441, Lunar and Planetary Institute, Houston, 1981).
26. Walker, G. P. L. *Phil. Trans. R. Soc.* **A274**, 107–118 (1973).

A model of the Antarctic Ice Sheet

J. Oerlemans

Institute for Meteorology and Oceanography, University of Utrecht, Princetonplein 5, Utrecht 3508 TA, The Netherlands

Numerical modelling of ice sheets and glaciers has become a useful tool in glaciological research. A model described here deals with the vertical mean ice velocity, is time dependent, computes bedrock adjustment and uses an empirical diagnostic relationship to derive the distribution of ice thickness in ice shelves. The rate of snowfall and ice/snow melt depends on the (prescribed) sea-level temperature, surface slope, elevation and distance to open water. The model is able to reproduce the major features of the Antarctic Ice Sheet. When it is run to a steady state for present climatic conditions, the main difference with the present ice sheet is that the shallow parts of the Weddell Sea become covered by grounded ice.

To understand ice-sheet histories from proxy data, and to see whether they are physically plausible, numerical ice-sheet models can have an important role. Although the theory of ice flow is formulated reasonably well (this does not apply to what happens at the ice–bedrock interface), modelling of large ice sheets requires a pragmatic approach. It is not possible to solve the full stress–strain rate relationships, together with the thermodynamic equation, for the whole Antarctic Ice Sheet, for example. This certainly applies if one wants to carry out integrations over 100,000 yr or so.

In recent years, ice-sheet models based on a flow law for the vertical mean ice velocity have become popular[1–4]; they have been used mainly for palaeoclimatic studies, and in complexity come between perfectly plastic models[5] and three-dimensional models[6]. In these models temperature is not calculated, so constant bulk flow parameters have to be used. In spite of the simplifications involved, the vertically integrated models capture most of the characteristics of large ice sheets.

Based on these observations, I have developed a vertically integrated model of the Antarctic Ice Sheet. In fact the Antarctic Ice Sheet is an excellent test case for the 'vertical mean approach', because it has a very irregular bedrock topography and subtle grounding line dynamics at some places, and its present-day physical characteristics are well documented[7].

The first version of the present model was used to investigate how sensitive the Antarctic Ice Sheet is to changes in the ice accumulation rate[2]. However, this version did not include bedrock adjustment and ice shelves, and could only be used for short integrations.

The second version, results of which are presented here, has much more internal freedom. The response of the bedrock topography to a varying ice load is computed, and a diagnostic relation for ice shelves (though simple) is included. A varying bedrock topography introduces the problem of initial conditions. It is not very well known how close the present bedrock topography is to isostatic equilibrium. A 100-m error in the equilibrium bedrock elevation will be insignificant in the central parts of the East Antarctic Ice Sheet, but is very important in such marginal regions as the Ross Sea.

Model description

The calculation of ice flow is based on a flow law relating vertical-mean horizontal ice velocity u to basal shear stress τ_b

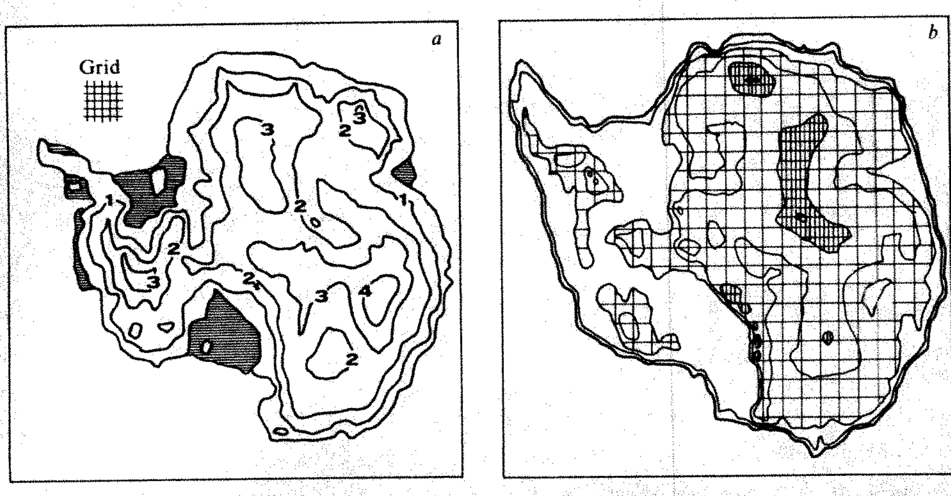


Fig. 1 *a*, Observed ice thickness as resolved on the model grid (part of which is shown in the upper left corner). Contour interval is 1 km. The major ice shelves are indicated by shading. *b*, Equilibrium bedrock topography in the case of no ice load. Contour interval is 1 km. The lowest contour level is 2 km below present sea level. Light shading indicates bedrock above sea level, heavy shading bedrock above the 2 km level. All input data are from ref. 10.

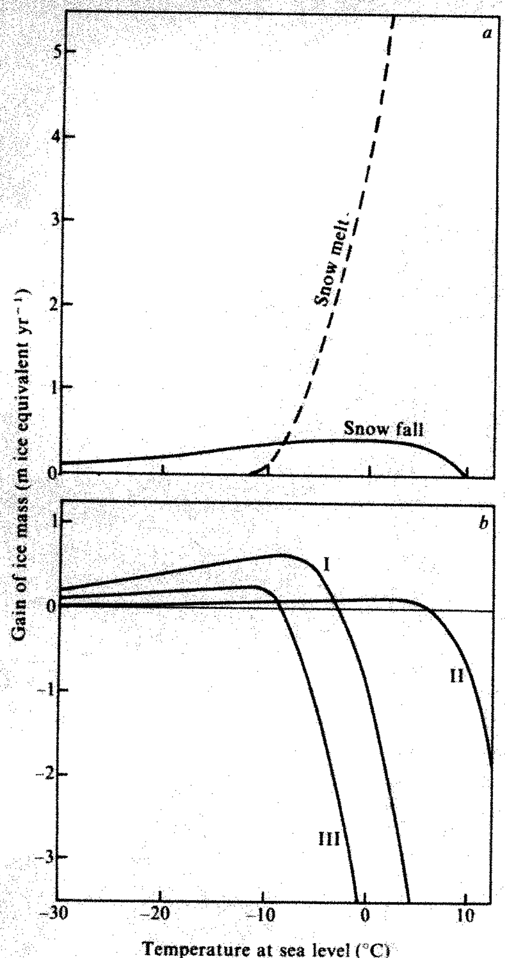


Fig. 2 Formulation of the mass balance, or ice accumulation rate, for the ice-sheet model. *a*, Dependence of snow/ice melt and snowfall on sea-level temperature for zero surface elevation and zero surface slope. *b*, The mass balance for different local conditions: I, a surface slope of 0.02 and elevation of 500 m, at the edge of the continent; II and III, the interior of the continent for elevations of 2,000 m and 0 m respectively.

(ref. 8). This law is of the Glen-type: $u = k\tau_b^m$, where k and m are flow parameters. The expression for u can be generalized to the two-dimensional case (involving assumptions that will not be discussed here), yielding the vertically integrated mass flow vector **M**:

$$\mathbf{M} = KH^{m+1}[\nabla H' \cdot \nabla H']^{(m-1)/2}\nabla H' \quad (1)$$

where H is ice thickness, ∇ is the two-dimensional (horizontal) gradient operator and $H' = H + h$ is the surface elevation (h is the bedrock elevation with respect to sea level). So ice flows down the surface slope $\nabla H'$, and the rate at which this happens increases strongly with both ice thickness and surface slope. The evolution of the ice sheet is determined by the conservation of ice mass, that is

$$\partial H/\partial t = -\nabla \cdot \mathbf{M} + G \quad (2)$$

G is the ice accumulation rate at the surface, expressed in units of ice depth per unit of time.

The response of the bedrock to the ice load is computed from:

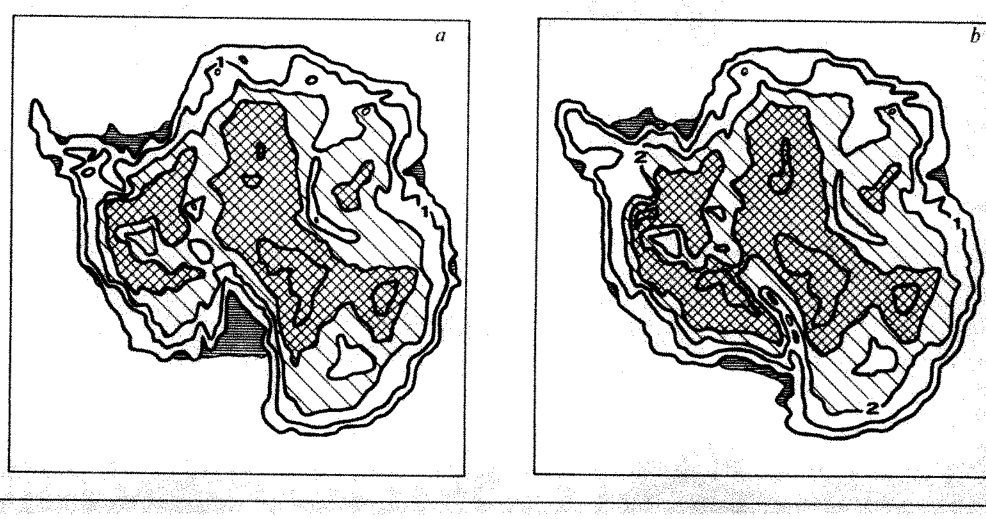
$$\partial h/\partial t = -(h - h_0 + H/3.2)/T \quad (3)$$

where h_0 is the equilibrium bedrock height without ice load. Equation (3) formulates a damped return, with time scale T, to local isostatic equilibrium. For loads with a horizontal length scale of more than a few hundred kilometres, equation (3) forms a good approximation[9]. For smaller loads, the flexural rigidity of the lithosphere becomes important. So, in general, equation (3) should work reasonably well, but for very small ice sheets or near the edge of a large ice sheet errors may occur (order of magnitude: 50 m).

Given the present bedrock topography and the distribution of ice thickness (Fig. 1*a*), and assuming that the present state of Antarctica is close to equilibrium, h_0 can be computed. Using data from the Russian Atlas of Antarctica[10], resolved on a grid with a spacing of 100 km, yields a no-ice bedrock topography h_0 as shown in Fig. 1b. This clearly shows how little of the bedrock beneath the present West Antarctic Ice Sheet is actually above sea level, even if the ice load is removed. Because we cannot be sure that the Antarctic Ice Sheet is close to a steady state, h_0 may be subject to errors. This applies in particular to the shallow Ross and Weddell seas, which were probably captured by grounded ice during the last ice age (20,000 yr ago). So it is possible that in these areas uplifting is still in progress[11]. Nevertheless, the bedrock topography displayed in Fig. 1*b* will serve as the initial condition for h, simply because a sound alternative is not available.

Ice shelves have to be taken into account to let the position of the grounding line be determined by the model. A prognostic equation for ice shelf growth/decay would not be feasible, because it would require a very small time step in the computations. Instead a diagnostic procedure is included. The thickness of floating ice at some point is calculated from the degree of enclosure (which is between 0° and 360°) by grounded ice, weighted with the distance to the grounding line and the ice thickness at the grounding line. In the present model only four directions (0°, 90°, 180° and 270°) are checked for each grid point to estimate the enclosure. An additional condition is that the ice shelf cannot be maintained if its thickness drops below 250 m. For an optimum choice of the proportionality constants appearing in this formulation, the calculated ice shelves for the

Fig. 3 Equilibrium states of the Antarctic Ice Sheet, computed with the numerical model for present sea level (*a*) and a sea-level stand of 100 m below the present one (*b*). In both integrations the initial condition was $H = 0$. Ice shelves are indicated by horizontal shading. Heavy lines are contours of ice thickness, except the outer one which gives the grounding line. The contour interval is 1 km. Light shading indicates ice thickness > 2 km, heavier shading ice thickness > 3 km.



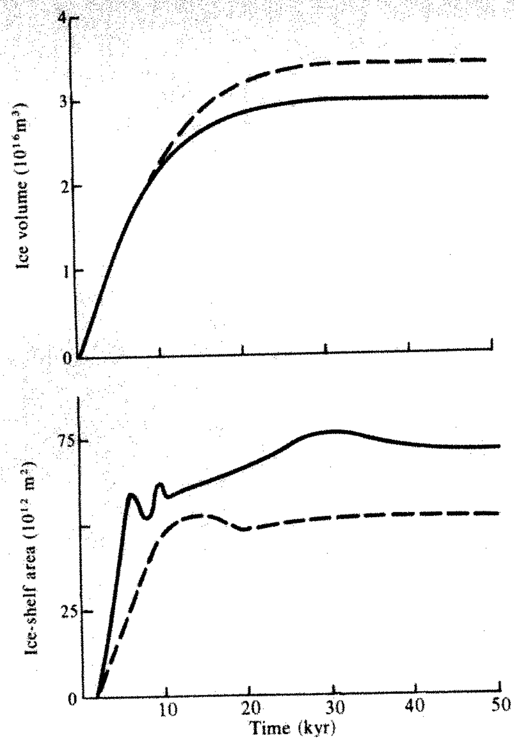


Fig. 4 Total ice volume and area covered by ice shelf, showing how the equilibrium states of Fig. 3 are reached. Solid line, experiment with present sea level; dashed line, experiment with the 100 m drop in sea level.

present distribution of grounded ice appear at the right locations with approximately the right thickness. Although ice-shelf dynamics does not explicitly occur in this procedure, it is dealt with in an implicit way, of course. It is no coincidence that ice shelves occur where the degree of enclosure is large. The compressive stresses associated with large enclosure counteract the ice thinning due to regular spreading. The present procedure mimics this mechanism well.

Because surface temperature and orographic enhancement of precipitation depend on the distribution of ice thickness, a general formulation for the ice accumulation rate G is desirable in which these effects appear in an explicit way. Based on the present distribution of surface elevation, surface slope and accumulation rate[10], the gain of ice mass due to precipitation is written as (in m ice equivalent yr^{-1}):

$$P = \max[0.05; f(\theta_0)(0.3 + 14\nabla H' - 4 * 10^{-5}H')/C] \quad (4)$$

So precipitation increases with surface slope and decreases with surface elevation. C is a 'continentality factor', which is computed by checking how many grid points in a circle of radius 500 km are covered by ice or bare ground. C ranges from 1 for an island of one grid point to 2 for the interior of a continent. The function $f(\theta_0)$, where θ_0 is sea-level temperature, models the dependence of snowfall on atmospheric water vapour content and the transition from snow to rain. Its shape is shown in Fig. 2a, in which a snowfall curve is drawn for $H' = 0$ and $\nabla H' = 0$. It is based on climatological practice and climatic change experiments with circulation models of the atmosphere[12,13].

Parameterization of ice/snow melt is based on mass-balance studies at the edge of the Greenland Ice Sheet[14]. With the assumption that the yearly cycle in surface temperature is more or less conserved in conditions of climatic change, the following approximate relation holds:

$$\begin{aligned} \text{melt} &= 0 && \text{if } \theta_s < -12\,^{\circ}\mathrm{C} \\ \text{melt} &= (12+\theta_s)^2 * 0.028 && \text{if } \theta_s \geq -12\,^{\circ}\mathrm{C} \end{aligned} \quad (5)$$

as for P, the unit is m ice equivalent yr^{-1}; θ_s is surface temperature which can be calculated from θ_0 by using a constant lapse rate γ (so $\theta_s = \theta_0 + \gamma H'$). Equation (5) only holds for temperatures below 0 °C or so, but this is sufficient because higher temperatures do not permit any ice cover to survive.

The difference between snowfall and ice/snow melt makes up the mass balance G. Given the surface slope and elevation, and the continentality factor, it only depends on sea-level temperature. The first three factors are determined by the shape of the ice sheet itself. So climatic change experiments can be performed by varying θ_0 only.

To illustrate how the mass balance depends on local conditions, Fig. 2b shows $G(\theta_0)$ for three typical conditions: (I) on the edge of an ice sheet, (II) in the interior of the continent at 2,000 m elevation, (III) in the interior at sea level.

The formulation of the mass balance described above is capable of reproducing all major characteristics of the present mass-balance field over the Antarctic Ice Sheet. It adds considerably to the internal freedom of the model. For example, zones of high precipitation rate follow the edge of the ice sheet, which is certainly realistic.

All model equations are solved with a numerical scheme on a grid with 100-km spacing between the grid points (see Fig. 1a). The time step used is 10 yr. Model parameters not yet specified are: $m = 2.5$ and $K = 0.5\ \mathrm{m}^{-3/2}\ \mathrm{yr}^{-1}$ (flow parameters), $T = 4{,}000$ yr (e-folding time scale for bedrock adjustment) and $\gamma = -8\,^{\circ}\mathrm{C\ km}^{-1}$ (lapse rate). Values of K and m were found by tuning the model to the present Antarctic Ice Sheet for prescribed and fixed bedrock elevation, ice accumulation rate and grounding line. However, the choice is not crucial: other combinations of m and K can be found to give similar results.

Results

In the first experiment the model was run to a steady state starting from no ice cover and bedrock topography h_0. The sea-level temperature was set at −14 °C. It takes ~30,000 yr before a steady state is approached, so the integration was

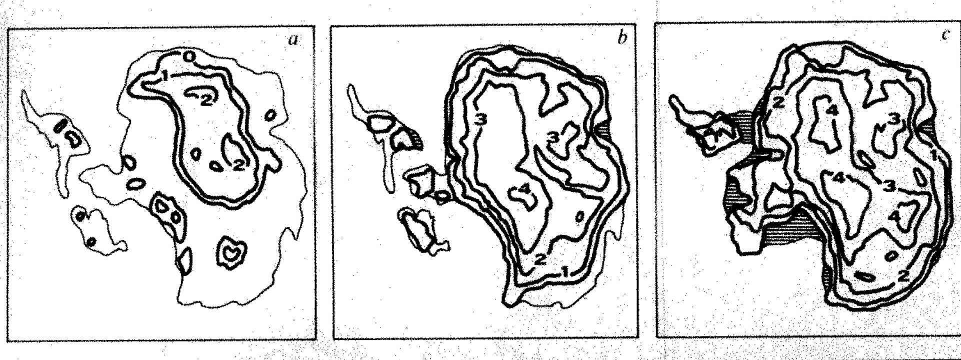


Fig. 5 Equilibrium states of the model Antarctic Ice Sheet for θ_0, the annual sea-level temperature: a, $\theta_0 = 5\,^{\circ}$C; b, $\theta_0 = 0\,^{\circ}$C; c, $\theta_0 = -5\,^{\circ}$C. Heavy lines are contours of ice thickness (km). Ice shelves are indicated by horizontal shading.

carried out over 50,000 yr This takes about 30 min central processor time on a CDC7600 computer

The calculated ice distribution is shown in Fig 3*a* Comparing this with Fig 1*a* shows that the model simulation is not too inaccurate, certainly not if one realizes that both bedrock elevations and mass balance are internally generated Also, note that the state of Fig 1*a* is not necessarily close to equilibrium, so that a sound judgement of the model performance cannot be made The major difference between the simulated equilibrium state and the present observed state of the Antarctic Ice Sheet is the extension of grounded ice in the vicinity of the Antarctic Peninsula In the model simulation a large part of the Weddell Sea is covered by grounded ice As a consequence, the ice thickness in the western and central parts of the Antarctic Ice Sheet is larger, and the total ice volume is ~25% more than the presently observed volume

The Ross Ice Shelf is well reproduced by the model, but the slope of the ice surface along the grounding line is steeper than observed This is probably because the model does not take into account enhanced ice-mass discharge by extensive basal sliding (which is supposed to operate in these regions)

Figure 3*b* also shows the equilibrium state calculated for a sea-level stand which is 100 m below the present one This was done to investigate the effect of glaciation in the Northern Hemisphere on the Antarctic ice volume The experiment was identical to the first one (the same initial conditions) except for the 100-m sea-level drop

Obviously, the West Antarctic Ice Sheet is sensitive to such a sea-level lowering the grounding line in the Ross Sea advances hundreds of kilometres At the upper (southern) tip of the Ross Sea, the ice thickness reaches 4 km Grounding line advance in the Weddell Sea is less spectacular, because in the present-day simulation a large part of the shallow sea is already covered by grounded ice Altogether, the volume of the West Antarctic Ice Sheet increases dramatically if the sea level lowers by 100 m According to Fig 4, the total ice volume increases by ~15% If sea level is reset to its present value, the ice sheet appears to remain stable Additional experiments revealed that a sea-level rise of at least 600 m is required to initiate grounding line instability in the Ross Sea (sea level temperature being equal) This suggests that in reality basal sliding must have had an important role in the deglacierization of the Ross Sea at the end of the last glacial extreme, because a 600 m sea-level rise did not occur, of course Basal melting may have been present during the grounding-line advance, thus keeping the ice thickness of the 'Ross Ice Sheet' small, or it may be triggered by the Holocene climatic warming However, such considerations are speculative

So far sea-level temperature θ_0 has been kept constant Note, however, how the model ice sheet reacts to changes in θ_0 Figure 5 shows steady-state ice sheets computed with $\theta_0 = 5, 0, -5$ °C respectively For $\theta_0 = 5$ °C, the mass balance is positive only in mountainous regions In general, the ice does not reach the coast but melts at the ice-sheet edge For $\theta_0 = 0$ °C, the ice sheet covers most of the eastern part of the Antarctic continent, but ice cover in the western part is still restricted to higher regions For $\theta_0 = -5$ °C, part of the West Antarctic Ice Sheet is formed, although the (present) Antarctic Peninsula is not yet attached to the main ice sheet A further drop of θ_0 of a few °C creates the ice sheet shown in Fig 3*a* There is not much in between a steady state in which the Weddell Sea is not covered by grounded ice did not appear

An interesting feature in these experiments is the large ice thickness occurring for $\theta_0 = -5$ °C in the central parts of East Antarctica (compare with Fig 3), this is due to the higher snowfall rates (see Fig 2) However, such findings critically depend on the parameterization of the mass balance and should be considered with caution

Discussion

The experiments described above should be considered as a first step in developing a more or less complete model of the Antarctic Ice Sheet Fine tuning of the model has not yet been attempted and several physical processes are not yet included (for example, the impact of extensive basal sliding) So what can we learn from these experiments?

First, it seems that a vertically integrated model is capable of reproducing the influence of mountain ranges on the distribution of ice thickness This is particularly clear if one considers an experiment in which the bedrock topography is fixed (see ref 2) In that case the model simulation is very good and errors are only due to the poor resolution of some bedrock irregularities In fact the value of the flow parameter *K* used in this study was found by tuning the fixed-bedrock model version to the present distribution of ice thickness Even if the mass balance and bedrock elevation are free, the simulated ice sheet closely resembles the present Antarctic Ice Sheet Moreover, because we do not know whether the ice cover in the Weddell Sea region represents an equilibrium state, we cannot state that the model result of Fig 3 is in error (there is some evidence that ice cover on the Antarctic Peninsula is expanding[15])

Experiments carried out by Budd and Smith[16] using a similar model of the Antarctic Ice Sheet produced results in agreement with those presented here, although they used a different flow law This again demonstrates that for large-scale modelling of ice sheets the precise form of the flow law is not important models can be easily tuned to give similar results

The Quaternary history of the Antarctic Ice Sheet[17] seems to fit in the picture emerging from the numerical experiments For a small sea-level drop (due to formation of ice sheets in the Northern Hemisphere) the ice sheet jumps to a state with large ice volume It returns to a state with small ice volume only if sea level and sea-level temperature are sufficiently high (during a Northern Hemisphere interglacial) However, the required increase in sea level and/or temperature is very large and additional mechanisms are apparently needed to remove the grounded ice from the Ross and Weddell seas Surging of the West Antarctic Ice Sheet could be one of them

Future development of this model will include a more refined treatment of ice shelves, computation of the temperature field within the ice and a calculation of the effect of basal sliding on ice-mass discharge In such a model ice surges may occur Experiments with a one-dimensional model version including thermodynamics have shown that such events can be modelled very well by vertically integrated models[18] The computational efficiency of the vertically integrated models (as compared with models with a grid in the vertical direction) will make it possible to study possible surging of the Antarctic Ice Sheet as a 'latitude–longitude problem' instead of looking along a steady flow line[19]

Received 8 February accepted 20 April 1982

1 Mahaffy, M A W *J geophys Res* **81**, 1059–1066 (1976)
2 Oerlemans, J *J Climatol* **2**, 1–11 (1982)
3 Birchfield, G E, Weertman, J & Lunde, A T *Quat Res* (in the press)
4 Pollard, D *Nature* **296**, 334–338 (1982)
5 Weertman, J *Nature*, **261**, 17–20 (1976)
6 Jenssen, D *J Glaciol* **18**, 373–390 (1977)
7 Budd, W F, Jenssen, D & Radok, U *Derived physical Characteristics of the Antarctic Ice Sheet* (Publ No 18 of Department of Meteorology, University of Melbourne 1970)
8 Nye, J F *J Glaciol* **3**, 493–507 (1959)
9 Turcotte, D L *Adv Geophys* 51–86 (1979)
10 *Atlas Antarktiki* (Glawnoje Oeprawlenie geodezii i Kartografii MG, 1966)
11 Greischar, L L & Bentley, C R *Nature* **283**, 651–654 (1980)
12 Manabe, S & Stouffer, R J *J geophys Res* **85**, 5529–5554 (1980)
13 Oerlemans, J & Vernekar A D *Contr atmos Phys* **54**, 352–361 (1981)
14 Ambach, W *Polarforschung* **42**, 18–24 (1972)
15 Sugden, D E & Clapperton C M *Nature* **286**, 378–381 (1980)
16 Budd, W F & Smith, I N *Annals Glaciol* (in the press)
17 Thomas, R H & Bentley, C R *Quat Res* **10**, 150–170 (1978)
18 Oerlemans, J *Free Oscillations of Polar Ice Sheets*, Preprint (University of Utrecht, 1982)
19 Budd, W F & McInnes, B J *Hydr Sci Bull* **24**, 95–104 (1979)

Effects of volatiles on mixing in calc-alkaline magma systems

Herbert E. Huppert*, R. Stephen J. Sparks† & J. Stewart Turner‡

* Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Silver Street, Cambridge CB3 9EW, UK
† Department of Earth Sciences, University of Cambridge, Downing Site, Cambridge CB2 3EQ, UK
‡ Research School of Earth Sciences, Australian National University, Canberra, Australia

The intimate mixing between different magmas of disparate densities, characteristic of some calc-alkaline magma systems, is explained by a new mechanism involving the emplacement of a layer of wet undersaturated mafic magma at the base of a magma chamber containing more differentiated magma Turbulent transfer of heat between the mafic magma and the overlying magma leads to crystallization and exsolution of volatiles in the lower layer For an initial water content of a few per cent the bulk density of the mafic magma can become equal to that of the overlying magma causing overturning and intimate mixing of the magmas The mechanism is most probable in the low pressure environment of a high level magma chamber

THE replenishment of a magma chamber by an influx of new magma from depth has been recognized as important in determining petrological and geochemical characteristics of volcanic and plutonic rocks[1] and in triggering volcanic eruptions[2–5] Previous investigations of the fluid dynamics of replenishment[3 6–9] have been directed mainly towards understanding systems involving dry basic magmas and have not been concerned with effects resulting from the release of gas In this article we investigate some of the fluid dynamical phenomena due to the presence and release of water vapour which can occur when a magma chamber is replenished by hydrous magma, and consider specifically magmas of the calc-alkaline association We note that the term 'calc-alkaline' is a common, but imprecisely defined way of describing magmas which are generally associated with orogenic belts and some island arcs Such magmas are usually, but not ubiquitously, characterized by low amounts of Fe enrichment and relatively high amounts of volatiles

There is increasing evidence that intimate mixing of magmas is a common phenomenon In basaltic magmas mixing can be explained in many circumstances because differentiated basaltic magmas can have the same density as higher temperature, more primitive basaltic magmas[3 6 9] Also the relatively low viscosity of many basaltic magmas allows rapid mixing between them However, evidence for magma mixing in volcanic rocks of the calc-alkaline association is also common The belief that mixing has an important role in the evolution of many basaltic andesites, andesites and dacites is based on observations of disequilibrium phenocryst assemblages, inhomogeneous glass compositions, wide ranges of compositions of glass inclusions in neighbouring phenocrysts, banded pumice fragments and mafic clots or micropillows incorporated in more silicic pumice or lava which show distinctive quench textures[5 10–14] Possibly the most convincing petrographical evidence for intimate mechanical mixing comes from the occurrence of bimodal phenocryst assemblages Many basaltic to andesitic lavas, for example, contain both calcium-rich and calcium-poor plagioclase phenocrysts and orthopyroxene and augite with high Fe/Mg ratio together with pyroxene with low Fe/Mg ratio[11–14] Petrological arguments often indicate that mixing between two or more compositionally distinct magma batches must have occurred only shortly before eruption, leading to the suggestion that mixing may even trigger an eruption[2 4 5 10] or alternatively that the mixing was a consequence of the eruption

If magma mixing is accepted as a widespread process in many calc-alkaline magma systems, there is a major physical problem to solve Dry magmas of the calc-alkaline association show a steady decrease in density with differentiation along with the decline in temperature[15] In a conventional model of a volcano, the plumbing system or chamber is filled with magma which gradually differentiates during repose periods If new undersaturated magma replenishes a chamber from depth (perhaps resulting in an eruption) and forms a lower layer it will, in general, be more primitive, hotter and denser But the decrease in density of the new magma due to crystallization and thermal effects alone is not sufficient to make the densities of the new and resident magmas equal Thus there does not at first sight appear to be a driving force in such a situation to cause mixing This argument, however, assumes that vapour is absent (no free volatiles) Calc-alkaline magmas are frequently rich in water, and Anderson[16] has summarized the evidence which suggests that the water content of many mafic to intermediate calc-alkaline magmas is often in the range 1–4% by weight Eichelberger[5] has shown that the presence of either dissolved or exsolved volatiles in calc-alkaline magmas has large effects on bulk density We continue along this line to demonstrate that even a small percentage of volatiles can have profound effects on the fluid dynamics of convection in replenished magma chambers

We consider in this article the influence of water exsolution on the densities of oversaturated hydrous mafic magmas and show that the induced changes in density can be substantial at low to moderate pressures This idea follows on from the study of Eichelberger[5] who, influenced by observed density differences of mafic inclusions within rhyolitic lavas, proposed that vesiculation of hydrous mafic magma at the interface between two layers can cause density changes large enough to allow mafic magma to become buoyant in more silicic magma However, only the relatively thin interfacial region would be destabilized by the process suggested by Eichelberger, and not the entire lower layer We aim to demonstrate that density changes in the whole layer of saturated hydrous basaltic or andesitic magma can be considerable and that, at low to moderate pressures, intimate mixing can take place between magmas of the calc-alkaline association

Effects of volatile release on magma density

It follows from basic principles that at equilibrium the bulk density of a hydrous saturated magma in the presence of excess water is given by

$$\sigma^{-1} = RT(N-n)/P + (1-N+n)/\rho \qquad (1)$$

where σ is the bulk density of the magma, R is the gas constant, T is the temperature in Kelvin, P is the pressure, N is the total weight fraction of water, n is the saturated weight fraction of water dissolved in the melt phase of the magma, and ρ is the

0028-0836/82/240554-04$01.00

weighted density of crystals plus melt The first term in equation (1) represents the contribution due to the gas and the second term that due to the crystals and melt Only if $N > n$ is there any gas, otherwise $\sigma = \rho$ The perfect gas law can be used in equation (1) because at the temperatures and pressures under consideration the departures from this law are less than a few per cent

The weight fraction n can be estimated from data on the saturation concentration of water in silicate melts[17] The solubility of water can be calculated from the empirical formula

$$n = s(1-x)P^{1/2} \tag{2}$$

where s is the solubility constant and x is the weight fraction of the crystalline phases in the magma

As an example, we consider the case of hydrous basaltic magma injected into the base of a magma chamber containing more fractionated, cooler and lower density magma (basaltic andesite, andesite or dacite) We make the simplifying assumptions that the crystallization of the basaltic magma involves only anhydrous phases The bulk density, ρ, of the residual melt and crystals has been calculated by assuming that the crystals have a constant density of 2 8 g cm^{-3} and by evaluating the decrease in density of the residual melt as water is concentrated by crystallization The density of the melt was calculated using the method of Bottinga and Weill[18] Variations of ρ with T alone due to thermal expansion are neglected because they are negligibly small in comparison with changes in bulk density of the magma due to volatile exsolution The solubility constant $s \approx 0\ 0014$ bar$^{-1/2}$ is taken from solubility data of water in Mount Hood andesite[17]

Finally, we require a relationship between temperature and crystal content The fractional crystallization of calc-alkaline magmas is quite complex and varies from volcano to volcano Thus a unique relationship cannot be accurately formulated However, inspection of estimates on the eruption temperatures and on fractional crystallization models of calc-alkaline magmas suggests that

$$T = 1{,}373 - 200x \tag{3}$$

is a plausible expression[14 19 20] The results of our density calculations are not in fact strongly dependent on temperature and thus an approximate relationship such as equation (3) seems acceptable

Figure 1 presents the results of bulk density calculations for a hydrous basalt crystallizing at 0 5, 1 5 and 3 0 kbar pressure with an initial temperature of 1,100 °C and zero initial crystal content The specific calculations presented are representative of the general behaviour of gas-rich magmas and are similar to the results of Eichelberger[5] From equation (2), the maximum total weight fraction of dissolved water in a silicate melt is represented by $N = SP^{1/2} \equiv N_0$ For $N < N_0$, the magma is initially undersaturated It continues to be undersaturated until the weight fraction of crystals $x = 1 - (N/N_0)$, when saturation first occurs During this stage the bulk density increases slightly as the proportion of crystals in the system increases Further crystallization involves exsolution of small water vapour bubbles and, assuming these remain suspended in the magma, leads to a continuously decreasing bulk density In general, if the total water content is decreased, keeping all other variables fixed, the density increases If the pressure is increased, keeping all other variables fixed, the density of the gas bubbles increases and thus so does the bulk density

The specific calculations demonstrate that the bulk density of hydrous basaltic magma can vary as a consequence of total water content and substantially as a result of volatile exsolution The approximate densities of dry degassed magmas of the calc-alkaline association are indicated on the right-hand side of Fig 1 It follows that at 1 5 kbar, for example, oversaturated basaltic magma can reach the same density as dry, or even volatile undersaturated, andesitic, dacitic or rhyolitic magma This comparison of hydrous basalt densities with 'dry' densities is merely presented to illustrate the very substantial variations that can occur in the bulk density of a fixed magma composition if excess volatiles are present in the lower layer

Even at pressures >1 5 kbar the changes can still be large enough to cause density inversions to occur in some magma chambers, although the total water contents required are at the high end of the range reported by Anderson[16] If water contents between 1 and 4% are typical of calc-alkaline mafic to intermediate magmas, the density changes necessary for convective mixing to take place can occur in shallow-level magma chambers

Geological applications

When mafic magma is injected into the base of a chamber containing less dense, more differentiated magma, any mixing

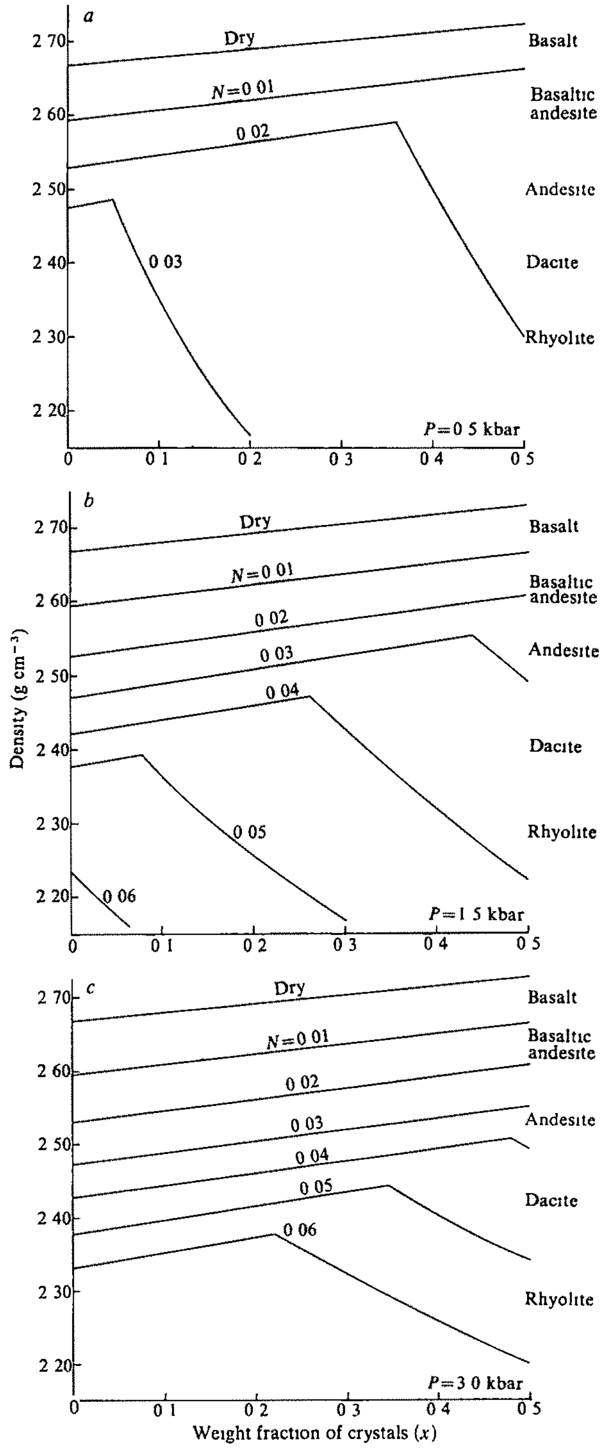


Fig. 1 The bulk density of wet basaltic magma as a function of the weight fraction of crystals for various total weight fractions of water, N, at a, 0 5 kbar, b, 1 5 kbar, and c, 3 0 kbar

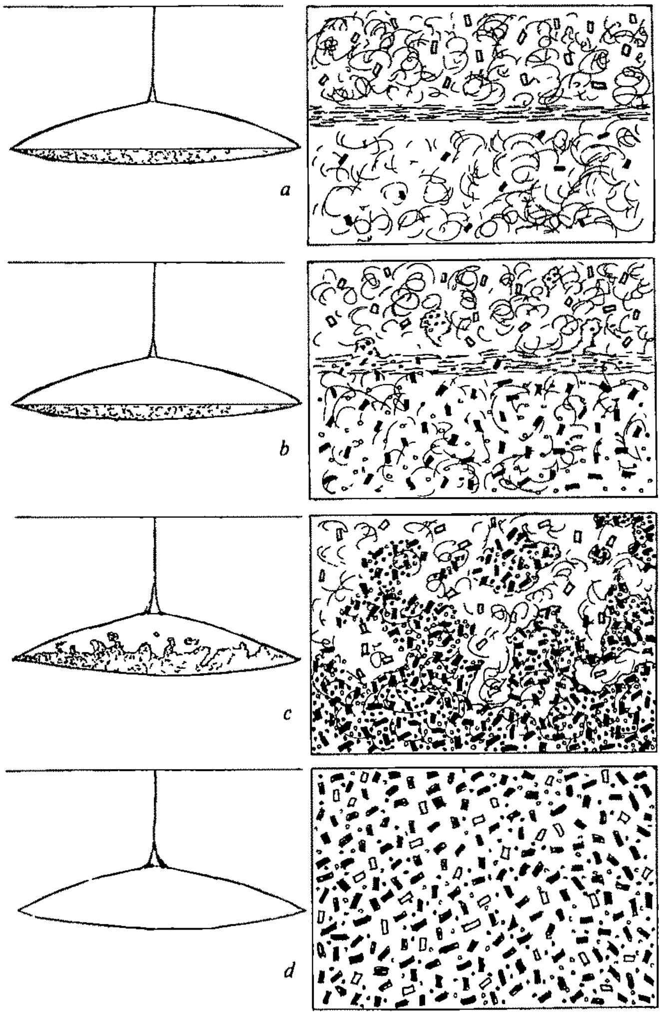


Fig. 2 Schematic model for a magma chamber replenished by an influx of hot, wet basaltic magma The four diagrams on the right represent close-up views at the height of the interface of the corresponding diagrams on the left *a*, The heavier magma of the lower layer loses heat across the interface and crystallizes *b*, The magma of the lower layer becomes saturated and further crystallization produces exsolution of water vapour bubbles *c*, The bulk density of the magma of the lower layer becomes equal to that of the overlying magma and the lower-layer magma rises into the upper layer *d*, Intimate mixing between the magmas results

will initially be strongly inhibited by the density contrast A two-layered or multi-layered system, with a horizontal interface separating the recently injected hot, denser, mafic magma from the cooler, less dense, silicic magma above, therefore seems a likely precursor to any well-mixed state Subsequent crystallization is one process for producing a decrease in density of the lower layer In the case of picritic basalt magmas, crystallization of dense components, notably olivine, leads to a decrease in the density of the residual liquid which can be sufficient to cause overturning and mixing with overlying differentiated basalt[6,7] In the case of calc-alkaline magmas, changes in the density of the liquid composition due to crystallization alone may not be sufficient to result in overturn We suggest that it is the release of water vapour once the lower layer has become saturated, which is essential

The overall process we envisage has a qualitative resemblance to one proposed[6,7] for basaltic magma chambers We suppose that hydrous mafic magma enters a chamber containing more differentiated magma, as depicted in Fig 2*a* This influx may be a consequence of tectonic activity and could possibly trigger an eruption[2,4] We assume that the volume of new magma forming the lower layer is much less than that of the upper layer and so the evolution of the lower layer proceeds with little change in the upper layer This is depicted quantitatively for the case of a dry basaltic magma chamber in Fig 7 of ref 7 As in our previous analyses of dry basaltic systems, the hydrous mafic magma transfers heat through a sharp interface and crystallizes (Fig 2*a*) As crystallization proceeds, the weight fraction of dissolved water increases in the residual liquid until the magma is saturated It can be shown[7,21,22] that the Rayleigh number of layers, even as thin as a few metres, of viscous andesitic magma at a temperature excess of 100 °C often exceeds 10^6 Thus the convective motions will be turbulent, as depicted in Fig 2

After the magma saturates, small bubbles form and, along with the crystals, are held in suspension by the vigorous turbulent convective velocities in the lower layer (Fig 2*b*) The bubbles remain in suspension because the velocity of rise of small gas bubbles relative to quiescent fluid is insignificant in comparison with the r m s turbulent vertical velocities in the lower layer Similar reasoning applies to the crystals, quantitative calculations for the velocities in this case are given in ref 7 The exsolution of water vapour causes the bulk density of the lower layer to decrease continually in accord with the results presented in Fig 1 Eventually, the bulk density may become equal to that in the overlying magma (which has changed by only an insignificant amount), whereupon overturning and hybridization of the magmas occur (Fig 2*c*, *d*) Further reactions involving the gas phase may be possible during the mixing We emphasize that we chose the simple and arbitrary chamber geometry depicted in Fig 2 for schematic purposes only The processes we have described are fundamental and would also occur in more complex and geometrically irregular plumbing systems Note that the magma in the upper layer need not be compositionally homogeneous, the same processes would occur if the upper layer were compositionally zoned, though the mixing would be confined to the lower parts of the system[9]

Eichelberger[5] proposes a different mechanism for mixing in which the interface between the mafic and silicic layers vesiculates and becomes unstable This mechanism seems possible and is incorporated in Fig 2*b* However, the interface thickness can be calculated[7] to be quite small, of the order of a few centimetres for temperature differences between layers of ~100 °C, and will thicken to only a few metres as the layers exchange heat Thus, while instability of a vesiculating interface may account for the occurrence of some mafic inclusions in silicic volcanic rocks, it cannot account for the mixing of large volumes of magma

There are several further predictions which follow from the general model when applied directly to calc-alkaline systems For a given pressure the mafic magma must have a water content within a restricted range for the mechanism to be viable For example, as illustrated in Fig 1*b*, at 1 5 kbar the total water content for the mafic magma to be initially undersaturated must be <5 4% and, assuming that a fluid model is applicable only up to a crystal content of 50%, the total water content for saturation to occur must be >2 8% Overturning into an upper rhyolitic layer of density 2 3 g cm^{-3} will not occur unless the total water content exceeds 3 8% If the upper layer of andesitic magma has degassed however, the total water content of the mafic magma needs only to exceed 3 0% Thus there will be situations in the calc-alkaline association where overturning cannot occur and a final compositional stratification will result Due to the variation of these limits on water content with chamber pressure, with the composition of the more differentiated magma and with the water content of the input mafic magma, both compositional zonation and intimate mixing could occur in the same magmatic system, but at different depths The data in Fig 1 suggest that the mechanism we propose is most easily achieved in high-level conduits and near surface chambers beneath volcanoes

There are considerable viscosity variations amongst calc-alkaline magmas, which will result in large variations in the time scale of the evolution of the layers A quantitative analysis can be made for the heat transfer in a two-layer system[7] and,

from equation (11a) of ref 7, the temperature and crystallization history of the lower layer can be predicted For basaltic systems the calculations suggest that the time taken to overturn can range from weeks to years depending on the initial thickness of the lower layer In calc-alkaline magmas, silica enrichment can result in increases in viscosity from two to more than four orders of magnitude over the viscosity of tholeiitic basalts Application of the two-layer system model[6,7] indicates that the time for overturn varies directly as the cube root of the viscosity, and hence overturn times may be increased by factors of order 10 or more for calc-alkaline magmas Incubation times are then typically years or decades For more complex layered systems, a similar overturn mechanism can occur[9], but the thermal history of the system cannot yet be quantified However, overturn times are likely to be considerably longer

In connection with volcanic activity, we note that in a rigid chamber of fixed volume, exsolution of gas in the lower layer must result in increasing chamber pressure Blake[4] has shown that an increase of chamber pressure sufficient to exceed the chamber wall fracture strength can occur by influx of new magma of volume typically of the order of 10^{-3} of the chamber volume Exsolution of a gas phase in the lower layer will result in increasing chamber pressure, and thus the generation of gas due to crystallization could trigger an eruption For example, if the lower layer constitutes 5% of the total chamber volume and a density decrease of 0 1 g cm^{-3} occurs due to vesiculation in this layer, the volume of gas will be 2×10^{-3} of the chamber volume Thus many situations can be envisaged where eruption could be triggered by vesiculation in the lower layer

We also note that the mixing during overturn may produce reactions between the magmas and volatiles which could be significant in driving or triggering volcanic eruptions Vesiculated mafic magma from the lower layer will be transported during the mixing process to environments which, on average, will be lower in temperature and pressure Further crystallization of the mafic magma during mixing will yield additional gas release The bubbles will have to expand, or alternatively increase their internal pressure, which will generate extra chamber pressure to drive magma to the surface

Most of the convincing petrographical evidence for hybrid magma in the calc-alkaline association comes from stratovolcanoes erupting small-to-medium volumes of mafic to intermediate volcanic products[10,12,14] In such systems stagnant magma occupies the high-level conduits and reservoir systems beneath the summit during quiescent periods Such magma will have undergone partial or complete degassing during previous periods of activity and periods of quiescence We envisage that this magma will become either just saturated in volatiles, or in some cases undersaturated if magma degassed at the surface sinks back into the shallow-level reservoir Many eruptions from such volcanoes are envisaged to occur during ascent of fresh magma from depth We suggest that as new magma arrives in the shallow-level system, oversaturation can occur and that mixing will ensue between the old degassed magma and the new volatile-bearing magma

There are other mechanisms which can produce magma-mixing phenomena For example, as shown in Fig 1, a saturated mafic magma containing excess volatiles can be lower in density than the more differentiated resident magma In this case the new magma is immediately buoyant and ascends as either a laminar or an entraining turbulent plume[3] Mixing can also occur by sidewall crystallization of a mafic magma forming a turbulent layer of light silicic melt[23,24] Mixing between the boundary-layer fluid and chamber fluid can form hybrid magma, although there is no evidence that this mechanism can produce large volumes of hybrid magma There are no doubt other situations which also lead to local mixing within the complex, near-surface plumbing systems of volcanoes We have tried to document a plausible mechanism which is fluid-dynamically possible, and also consistent with conventional ideas on the plumbing systems of volcanoes The mechanism is capable of explaining the common occurrence of hybrid calc-alkaline volcanic rocks In some circumstances basaltic magma can become lighter than even dry rhyolitic magma and thus mixing of mafic magmas derived from the mantle and silicic magmas derived from the crust could occur

We thank A T Anderson, R J Arculus, J C Eichelberger and B D Marsh for helpful comments on a previous draft of the manuscript The paper is contribution 228 from the Department of Earth Sciences

Received 9 February, accepted 30 April 1982

1 O'Hara, M J & Matthews, R E *J geol Soc* **138**, 237–277 (1981)
2 Sparks, R S J *et al Nature* **267**, 315–318 (1977)
3 Sparks, R S J *et al Earth planet Sci Lett* **46**, 419–430 (1980)
4 Blake, S *Nature* **289**, 783–785 (1981)
5 Eichelberger, J C *Nature* **288**, 446–450 (1980)
6 Huppert, H E & Sparks, R S J *Nature* **286**, 46–48 (1980)
7 Huppert, H E & Sparks, R S J *Contr Miner Petrol* **75**, 279–289 (1980)
8 Huppert, H E & Turner, J S *Earth planet Sci Lett* **54**, 144–152 (1981)
9 Huppert, H E *et al Earth planet Sci Lett* **57**, 345–357 (1982)
10 Anderson, A T *J Volcan geotherm Res* **1**, 1–33 (1976)
11 Eichelberger, J C *Nature* **275**, 21–27 (1978)
12 Luhr, J F & Carmichael, I S E *Contr Miner Petrol* **71**, 343–372 (1980)
13 Sakuyama, M *J Volcan geotherm Res* **5**, 179–208 (1979)
14 Sakuyama, M *J Petrol* **22**, 553–583 (1981)
15 Bottinga, Y & Weill, D F *Am J Sci* **269**, 169–182 (1970)
16 Anderson, A T *J Geol* **87**, 509–531 (1979)
17 Burnham, C W *Geochim cosmochim Acta* **39**, 1077–1084 (1975)
18 Bottinga, Y & Weill, D F *Am J Sci* **269**, 169–182 (1970)
19 Rose, W I *et al J Geol* **85**, 63–87 (1977)
20 Arculus, R J & Wills, K J A *J Petrol* **21**, 743–799 (1980)
21 Bartlett, R W *Am J Sci* **267**, 1067–1082 (1969)
22 Rice, A *J geophys Res* **86**, 405–417 (1981)
23 McBirney, A R *J Volcan geotherm Res* **7**, 357–371 (1980)
24 Turner, J S & Gustafson, L B *J Volcan geotherm Res* **11**, 93–125 (1981)

Population dynamics of human helminth infections: control by chemotherapy

Roy M. Anderson* **& Robert M. May†**

* Department of Pure and Applied Biology, Imperial College, London University, London SW7 2BB, UK
† Department of Biology, Princeton University, Princeton, New Jersey 08544, USA

An analysis is presented of the population dynamics of the major helminth parasites of man with the aim of understanding observed patterns in the age-specific prevalence and intensity of infection Mathematical models are used to investigate the possibility of controlling helminth diseases by mass chemotherapy, and to explore the advantages of selective treatment of the most heavily infected individuals in a community

OVER 30 years ago Stoll[1] noted that the major helminth infections of man were among the most prevalent of all human infectious diseases Today the global picture is similar, and we still live in a very 'wormy world' there are roughly 1,000 million cases each of roundworm (*Ascaris lumbricoides*), pinworm (*Enterobius vermicularis*) and whipworm (*Trichuris trichuria*), over 600 million cases of hookworm (*Necator americanus* and *Ancylostoma duodenale*), 300 million cases of filarial infections

0028-0836/82/240557—07$01 00

(principally *Wuchereria bancrofti* and *Onchocerca volvulus*) and over 200 million cases (500 million people at risk) of schistosomiasis (*Schistosoma mansoni, Schistosoma haematobium, Schistosoma japonicum* and *Schistosoma intercalatum*)[2-5]

In contrast to most viral and bacterial diseases, helminth populations exhibit a remarkable degree of temporal stability in human communities[6,7] These parasite populations tend to be robust to perturbation, resulting from either seasonal changes in climate or the introduction of control measures This is in part a consequence of the human host's inability to acquire strong immunity to re-infection, so that helminth infections are persistent in character with hosts being continually re-infected In endemic regions of the world, a newborn child can expect to harbour worms for most of its life (Fig 1)

The wide range of methods used in attempts to control helminth infections include chemotherapy, improved sanitation and hygiene, education and (in the case of indirectly transmitted infections such as schistosomiasis and filariasis) vector control[8-10] New techniques, based on the use of monoclonal antibodies, raise the exciting possibility that vaccination against human helminths may be feasible in the future[11] Today, safe, effective and cheap chemotherapeutic agents are available for many of the major helminth infections of man (excluding the filarial worms)[12,13] There remains, however, a need for a better understanding of the most cost-effective ways of delivering these drugs, not at the individual level, but in the community as a whole Among the unresolved epidemiological questions concerning the use of antihelmintic agents are the following What proportion of the population must be treated, and how often, to achieve a defined reduction in parasite prevalence and abundance? How rapidly will the parasite population return to its pre-control level once chemotherapy ceases? Is the selective treatment of the more heavily infected individuals of greater benefit to the community, both in terms of costs and of reducing the prevalence of disease symptoms, when compared with random or blanket chemotherapy? This article examines certain theoretical and observational aspects of these questions

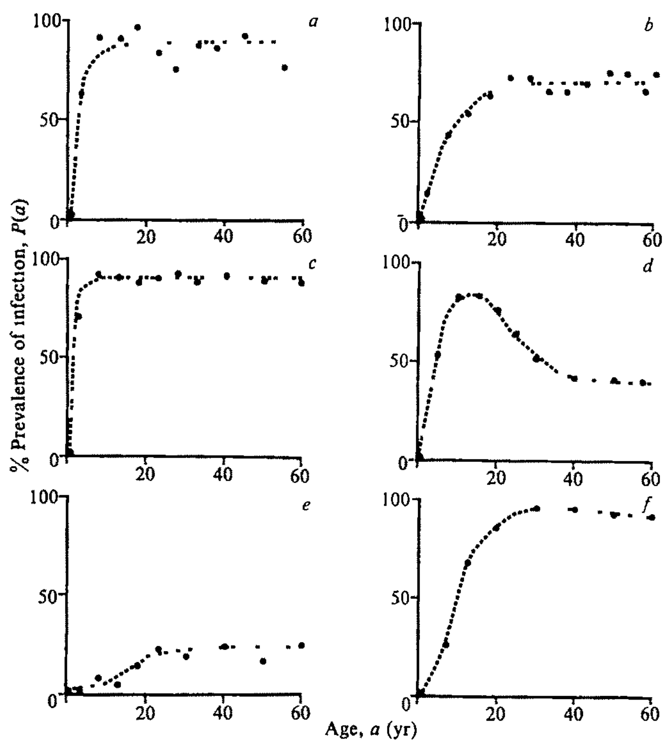


Fig. 1 Examples of horizontal age–prevalence data for some major helminth parasites of man in communities where the infections are endemic *a*, *A lumbricoides* in Iran[16], *b*, hookworm (both *Necator* and *Ancylostoma*) in Taiwan[60], *c*, *T trichuria* in the Philippines[61], *d*, *S haematobium* in Tanzania[34], *e*, *W bancrofti* in India[62] (prevalence of microfilaria carriers), *f*, *O volvulus* in Mali[49] The dashed lines are fitted by eye to illustrate the general trends, the solid dots are observed values

Population dynamics

An understanding of the factors which regulate parasite abundance, and contribute to the observed stability of helminth populations, is of obvious significance to the design of optimal control policies[14,15] Recent attempts to extend the pioneering work of Macdonald can be summarized as identifying five principal factors controlling the observed dynamics of the major helminth infections of man[7,16-18]

(1) The relative time scales on which the dynamics of the host and parasite populations operate These are largely determined by the expected life spans of the human host and the various developmental stages of the parasite The human host typically has a life span an order of magnitude or more in excess of any of the parasitic stages (Table 1*a*) For example, the adult worms in the human host have life spans of 1–16 yr depending on the species of parasite Note, however, that measurement of this parameter is difficult in practice and hence the values in Table 1*a* are only approximations Free-living infective stages plus infected vectors have much shorter life expectancies than those of the adult parasites These empirical observations enable certain simplifications to be made in the analysis of parasite population behaviour, without any serious loss of detail or accuracy[7,15-17] Given the information in Table 1*a*, we can examine the dynamics of the adult parasites under the assumptions that the host population is effectively constant in size on a time scale appropriate to changes in adult parasite population size and that the populations of the infective stages (or vectors) are essentially at equilibrium due to the rapidity with which changes in these populations occur when compared with those in the adult worm populations[14,15]

(2) The regulation of parasite population growth in the human host Although man is unable to develop an effective protective immunity to re-infection, he can mount immunological and nonspecific responses to the invasion of helminth parasites, which act in conjunction with resource limitation to reduce worm survival, establishment and fecundity in a manner that depends on the density of parasites in the host (refs 19, 20, Fig 2) These density-dependent responses are the major regulatory mechanisms controlling the growth of the total parasite population in any human community The net severity of these constraints in the host population depends critically on the statistical distribution of worm numbers per host[21-23]

(3) Worm distributions these are typically highly aggregated[21,22,24], with the majority of worms being harboured by a minority of the human population The generative mechanisms of such patterns are many and varied but include genetic, spatial and behavioural factors The negative binomial probability distribution has proved to be a good empirical model for observed patterns, this distribution is defined by two parameters, the mean worm burden per host, M, and a parameter, k, which varies inversely with the degree of parasite aggregation Observed k values are very small, and it is not unusual for >65% of the worm population to be harboured by <15% of the human community (Table 1*b*) This high degree of contagion enhances the severity of density-dependent constraints on parasite population growth[23]

(4) The reproductive biology of the parasite The major human helminths are dioecious and the production of transmission stages therefore depends on the probability that a female worm is mated This probability is a function of the statistical distribution of parasite numbers per host, there exists a critical average worm burden, termed the breakpoint, below which mating frequency is too low to maintain the parasite[7,17,25,26] The precise level of this breakpoint is in part determined by the degree of worm contagion in the human population Recent studies of *Ascaris* and hookworm infections[7,16], both of whose distributions tend to be highly aggregated in human populations, suggest that the breakpoint is typically very low—of the order of 0 3–0 5 worms per host Mating prob-

abilities thus seem to be of limited significance to the overall population stability of helminth parasites, and the breakpoint concept therefore appears to have little practical significance for disease control[7 16 26]

(5) The reproductive or transmission potential of the parasite The basic reproductive rate, R_0, is used to measure this potential and is defined as the average number of female offspring produced throughout the lifetime of a mature female parasite which themselves achieve reproductive maturity, in the absence of density-dependent constraints[7 16 25 27] If the parasite population reaches a steady state (stable endemic disease) the effective reproductive rate, R, is equal to unity, in other words, each female parasite exactly replaces herself in the next generation R_0 has recently been estimated for several helminth infections in different regions of the world[7 16 28], these estimates typically lie in the range 1–5 and are generally lower than those recorded for many viral, bacterial and protozoan infections of man[29 30 70] (Table 1c) The condition $R_0 = 1$ defines a transmission threshold below which the parasite is unable to maintain itself in the human population[7 16 25] The value of R_0 is determined by many individual components representing the birth, death and transmission rates of the various developmental stages in the parasites' life cycle Estimates of R_0, however, can be obtained directly from epidemiological data recording the prevalence and intensity of infection in different age classes of the host population The parameter R_0, in conjunction with the life expectancy of the adult parasite, A, the degree of worm aggregation, k, and the severity of density-dependent constraints on worm population growth, z, acts to determine the average intensity, M^*, and prevalence, P^*, of infection at equilibrium in a given community

Mathematical models

It follows from the relative time scales recorded in Table 1a that a description of the temporal dynamics of parasite population growth can be given in terms of one variable, $M(t)$, the mean adult worm abundance per host at time t[7 14 15] Assuming the parasites to be distributed in a negative binomial manner in the human population, with clumping parameter, k, the dynamics of infection by helminths in any defined human community may be described by a single differential equation for $M(t)$, of the general form

$$\mathrm{d}M/\mathrm{d}t = (M/A)\,[R_0 f(M,k,z) - 1] \qquad (1)$$

R_0, A and k are as defined above, z is an inverse measure of the severity of density dependence (see Fig 2) while f denotes the net force of these constraints acting on parasite population growth in an individual host For example, in ascariasis, density-dependent processes seem to act mainly on worm fecundity[16], such that, within a given host, the rate of egg production per female worm decays approximately exponentially as worm burden increases (Fig 2) This gives

$$f(M,k,z) = [1 + (1-z)M/k]^{-(k+1)} \qquad (2)$$

and the equilibrium average worm burden, M^*, and prevalence, P^*, are given by

$$M^* = [R_0^{1/(k+1)} - 1]\,[k/(1-z)] \qquad (3)$$

and

$$P^* = 1 - [1 + M^*/k]^{-k} \qquad (4)$$

If density-dependent constraints severely restrict worm fecundity the hosts with the maximum egg output may not

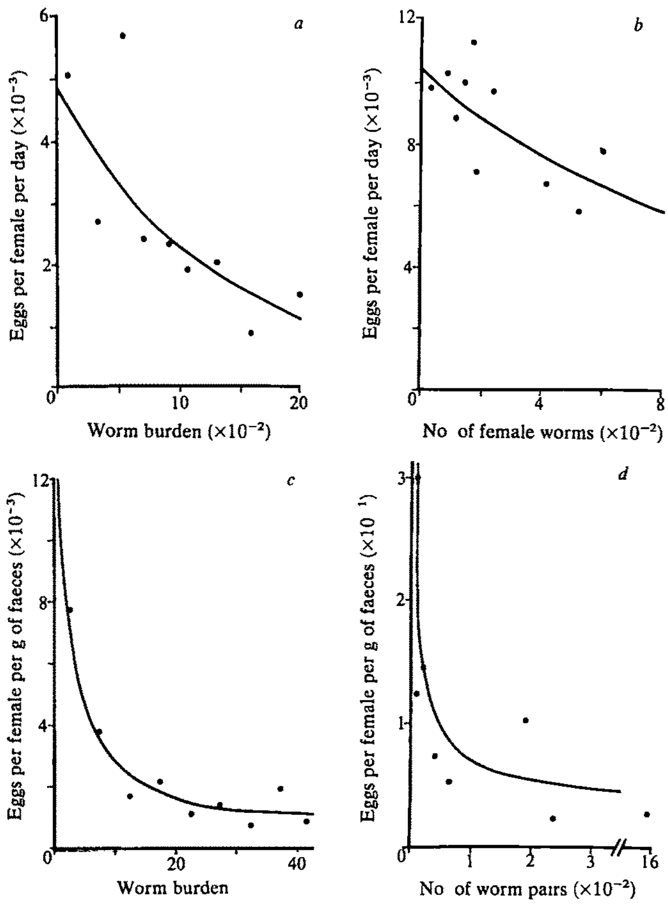


Fig 2 Examples of the decline in worm fecundity (in terms of egg output) as the burden of parasites in the human host increases (density-dependent egg production) ●, Observed values, —, best-fit power or exponential models a, *N americanus* (data from ref 63) The solid line denotes the best-fit exponential model of the form $y = \hat{a}\,e^{-\hat{b}i}$ where y is the egg output per female worm per day, i is worm burden per host, and $\hat{a}$ and $\hat{b}$ are constants with values $\hat{a} = 4\ 8\times10^3$, $\hat{b} = 0\ 73\times10^{-3}$ ($r^2 = 0\ 7$) b, Mixed infections of *N americanus* and *A duodenale* (data from ref 64) The solid line denotes the best-fit exponential model, as defined for a, with parameter values $\hat{a} = 10\ 3\times10^3$, $\hat{b} = 0\ 7\times10^{-3}$ ($r^2 = 0\ 7$) c, *A lumbricoides* (data from ref 16) The solid line denotes the best-fit power model of the form $y = \bar{a}i^{\bar{b}}$ where y and i are as defined for a, and $\bar{a}$ and $\bar{b}$ are constants with values $\bar{a} = 9\ 0\times10^{-3}$, $\bar{b} = -0\ 6$ ($r^2 = 0\ 9$) d, *S mansoni* (data from refs 59, 65) The solid lines denote the best-fit power model as defined for c with parameter values $\bar{a} = 38\ 4$, $\bar{b} = -0\ 36$ ($r^2 = 0\ 7$)

Table 1 Characteristics of the population biologies of human helminth parasites

a, Time scales

Parasite	Life expectancy: Man (yr)	Adult parasite in man (yr)	Infective stage or intermediate host	Ref
A lumbricoides	40–60	1–2	1–3 months (egg)	16
T trichuria	40–60	2–3	1–2 months (egg)	8
N americanus	40–60	2–4	3–10 days (L_3 larvae)	7
S mansoni	40–60	3–4	3–6 weeks (infected snail)	55, 56
O volvulus	40–60	8–14	1–3 weeks (arthropod vector)	8, 28, 57
W bancrofti	40–60	12–16	1–2 weeks (arthropod vector)	8, 58

b, Degree of aggregation of adult parasites in the human population, k

Parasite	Geographical location	k*	Ref
A lumbricoides	Iran	0 2–2 9	16
N americanus	India	0 03–0 6	7
N americanus and *A duodenale*	Taiwan	0 05–0 4	7
S mansoni	Brazil	0 03–0 5	59

c, Basic reproductive rate, R_0

Parasite	Geographical location	R_0	Ref
O volvulus	Sudan	5–35	28
A lumbricoides	Iran	4–5	16
N americanus	India	2–3	7
S mansoni	Brazil	1–2	59

* (k, An inverse measure of aggregation)

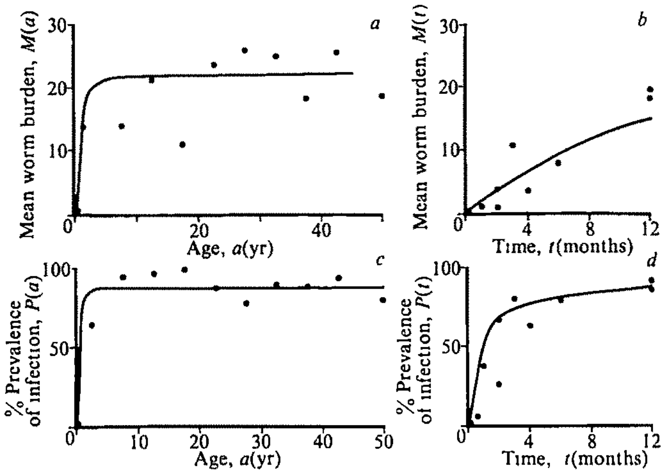


Fig. 3 The fit of the model defined by equations (8) and (9) in the text to data of the epidemiology of *A lumbricoides* in Iran (data from ref 16). ●, Observed values, —, model predictions. *a*, *c*, Rise in the average worm burden and prevalence of infection, respectively, with host age. The data were collected by means of a horizontal epidemiological survey[16]. *b*, *d* Show, respectively, the increase in parasite intensity and prevalence with time in a sample of people who were cleared of worms by antihelmintic treatment at time $t=0$. Note the rapidity with which both the average worm burden and prevalence approach their pre-control levels (indicated by the plateau levels in *a*, *c*). Estimates of the parameters of the model [equations (8), (9)] are as follows $R_0=4.3$, $M^*=22$, $A=1$ y, $z=0.96$, $k=0.57$. The density-dependent function, f, used in the model was as defined in equation (2) and the parameter values listed above were used to make the predictions in all four graphs, *a–d*.

necessarily harbour the maximum parasite burden. However, given the typical maximum worm burdens recorded in human communities, the observed patterns of density-dependent fecundity (Fig 2) suggest that such situations are unlikely to arise in practice. The qualitative dynamical behaviour of the model defined in equation (1) is independent of the precise form of function f. The quantitative properties, however, will depend on whether density-dependent processes act on worm fecundity, survival or establishment, or any combination of these. Human helminth fecundity is invariably density dependent (see Fig 2) but information on survival and establishment is extremely limited, and practically difficult to obtain.

Equation (1) may be used to approximate the dynamics not only of directly transmitted infections (such as *Ascaris*), but also of indirectly transmitted ones (such as *Schistosoma*) since the parameter R_0 represents the appropriate combination of transmission factors for the complete life cycle of any given parasite species[7,14,15,28,31]. The epidemiological data most commonly available are from horizontal and longitudinal studies of age–prevalence and age–intensity relationships[32]. The model can be extended to describe the way the mean worm burden, $M(t,a)$, changes with time, (longitudinal changes), and host age, a (horizontal changes)[28]

$$\partial M/\partial t+\partial M/\partial a=\Lambda(t)-M/A \tag{5}$$

The function $\Lambda(t)$ is defined for notational convenience as

$$\Lambda(t)=(R_0/AL)\int_0^{\infty} l(a)\,M(t,a)\,f(M,k,z)\,\mathrm{d}a \tag{6}$$

where $l(a)$ is the probability that a host survives to age a and L denotes human life expectancy ($L=\int_0^\infty l(a)\,\mathrm{d}a$). For human communities in which the prevalence and average intensity of infection have remained relatively constant with time (where $\partial M/\partial t\simeq 0$), equation (5) reduces to a simple differential equation for the change in average worm burden, $M(a)$, with host age, in this equation, Λ is a constant defined by equation (6) with $M(t,a)$ replaced by $M(a)$. For endemic infections of *Ascaris* and hookworm, the average worm burden rises rapidly as host age increases to reach a plateau such that the mean worm burden, $M(a)$, in the majority of age classes is essentially equal to the mean worm burden of the total population [the M^* of equation (3)], consequently, a good estimate of Λ is given by

$$\Lambda=(R_0M^*/A)\,f(M^*,k,z) \tag{7}$$

It is here assumed that the distribution of parasites in the entire community (consisting of all age classes) remains approximately negative binomial in form with mean M^* and clumping parameter k. The precise overall distribution will generally be formed from a mixture of negative binomial distributions (one for each age class) as a consequence of age-dependent human mortality and differences in exposure to infection[33]. The use of a common aggregation parameter for the entire community, however, remains a good approximation provided parasite life expectancy is short in relation to that of man[33]. Available empirical evidence suggests that the negative binomial distribution remains a reasonable empirical model of hookworm and roundworm numbers per host even when a sample consists of a mixture of individuals from different host age classes[7,16,34].

In this limiting case of steady infection, when $\partial M(t,0)/\partial t\simeq 0$, the appropriately simplified form of equation (5) yields an age–intensity relationship of the form

$$M(a)=\Lambda A\,[1-e^{-a/A}] \tag{8}$$

Here Λ is given by equation (7). The predicted age–prevalence curve (the proportion of each age class infected) then follows as

$$P(a)=1-[1+M(a)/k]^{-k} \tag{9}$$

Equations (8) and (9) can be used to obtain estimates of Λ, A and R [using equation (7)] from observed age–prevalence and age–intensity data[7,16,28,35].

We have assumed that the net rate of infection, Λ, is independent of host age and depends only on the average number of transmission stages present in the host's environments. This is often the case for infections such as *Ascaris*[14], but in many instances, Λ varies with age as a result of changes in the contact rate with infective stages (or infected vectors), the severity of density-dependent factors (mediated by age-related changes in host resistance to infection) or the degree of worm contagion in different age classes of the population[7,28,31]. For schistosome and filarial infections, age-related changes in acquired resistance and contact with infection are known to be important, although much controversy still surrounds their relative significance[28,36] (Fig 1).

The model defined in equations (8) and (9) can provide an accurate description of observed trends[16]. Figure 3 presents data pertaining to the epidemiology of *Ascaris* in Iran. The model can be seen to mimic changes both in the observed horizontal trends of intensity and prevalence in different age classes of the population (Fig 3*a*, *b*) and in the re-infection with time of individuals who were given a single treatment with a chemotherapeutic agent.

Control by chemotherapy

Effective and safe antihelmintic agents are available for the treatment of most major helminth infections of man[12,37,38], including mebendazole[5,39,40,66,67] (for enterobiasis, ascariasis and trichuriasis), pyrantel pamoate[8] (for ascariasis, enterobiasis, trichuriasis and hookworm), levamisole[40,41] (for ascariasis), bephenium hydroxynaphthoate[8] (for *Ancylostoma*) and the recently developed praziquantel[42–45] (for schistosomiasis). However, questions concerning the best way to use these drugs for the benefit of the community rather than the individual remain to be answered[10]. The above models can be adapted to examine these issues: we focus first on the long-term effects of the use of mass chemotherapy and, second, on the selective treatment of the most heavily infected individuals in a community.

Mass chemotherapy in the long term. When a drug is administered randomly within a population by treating a pro-

portion, g, of the community per unit of time and the drug has an efficacy h (where h represents the average proportion of the worm burden killed by a single, or short course of treatment), the resulting increase in the per capita death rate of adult parasitic worms, c, is then[7,16]

$$c = -\ln(1-gh) \qquad (10)$$

By subtracting the net death rate due to treatment, cM, from the right-hand side of equation (1), the average worm burden, $\bar{M}$, at the new equilibrium established by the control programme can be shown to be

$$\bar{M} = k([R_0/[1-A\ln(1-gh)]]^{1/(k+1)}-1)/(1-z) \qquad (11)$$

It is assumed that density-dependent constraints act in the manner defined in equation (2); other assumptions give similar results. To eradicate the infection the effective reproductive rate R (the basic reproductive rate, R_0, modified by the action of chemotherapy) must be reduced below unity. To achieve this, the proportion treated per unit of time must be greater than a critical value, $\hat{g}$ (ref. 7):

$$\hat{g} = (1-\exp[(1-R_0)/A])/h \qquad (12)$$

Even when $g > \hat{g}$ treatment must be continued for a period of time greater than the maximum life span of the adult worm. If treatment ceases before the average worm burden is below the breakpoint (which, unfortunately, we have seen is usually close to zero worms per host), the parasite population will rapidly return to its pre-control level (the speed of return depends on the values of R_0 and A). For helminths, however, eradication is not always a practical goal, and control policies are more often oriented to the elimination of disease rather than of infection because the two are not necessarily equivalent[3,5,45,46,68].

The model defined in equation (5) can be used to examine the impact of a defined programme of chemotherapy on the prevalence and intensity of infection both with time and in different age classes of the population (Fig. 4). The numerical results shown in Fig. 4 mimic the dynamics of endemic *Ascaris* and illustrate two important points—the rapidity with which certain helminth infections return to their pre-control level once chemotherapy ceases[47], and the differential impact of antihelmintic treatment on the prevalence and intensity of infection. Simply monitoring prevalence may lead to the incorect conclusion that control measures have little impact on parasite abundance. The principal advantage of these models is that (given estimates of parameters R_0, A, k and z) they can be used to predict quantitatively the outcome of a defined programme of chemotherapy and to define the level of treatment required for eradication.

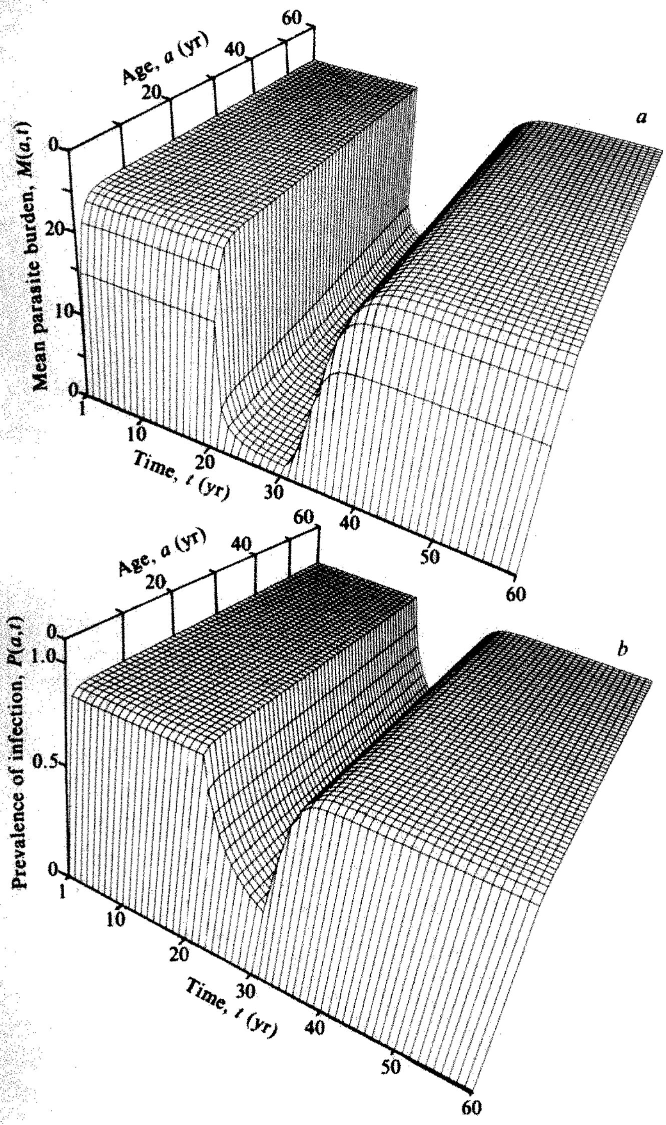


Fig. 4 Two three-dimensional graphs obtained by the numerical solution of the partial differential equation defined in equation (5). The simulations are designed to mimic the dynamics of endemic ascarasis in a human community, both with time and in different age classes of the population. *a*, Changes in the average worm burden; *b*, changes in prevalence. For the first 10 yr the helminth infection is at its endemic equilibrium (with M^* in the older age classes set at 22 worms per host). In year 10, mass chemotherapy is initiated and applied randomly in the community on a continual basis where 20% of the population is treated each month ($g = 0.25$ per month) with a drug of 95% efficacy ($h = 0.95$). After 10 yr the treatment is stopped and the graphs display the return of the parasite population to its pre-control equilibrium level. The human survivorship function, $l(a)$, was chosen to mimic typical patterns in the developing world with a life expectancy, L, of 50 yr. The density-dependent function f was as defined in equation (2) and the remaining parameters of the model were set at the following values: $R_0 = 4.3$, $k = 0.57$, $z = 0.96$, $A = 1$ yr. Note the rapidity with which the infection returns to its pre-control level once chemotherapy ceases and the differential impact of control on the average worm burden, $M(t,a)$, and the prevalence of infection, $P(t,a)$.

Single applications of mass chemotherapy. Antihelmintic drugs are often used intensively over short periods of time to reduce the prevalence of disease symptoms in a community. Frequently, such applications are given at irregular intervals with long time periods between successive treatments. The objectives are to maximize the short-term benefits and minimize the number of treatments (and hence cost). In 1924, Smillie[48] argued that the optimum way to apply mass chemotherapy was selectively to treat the most heavily infected individuals; Warren has recently re-opened this debate[5,49–51,69]. A major disadvantage of this strategy can be the extra cost involved in identifying the most heavily infected individuals; in practical terms this can be done either by faecal egg, urine egg or blood larval counts, or by noting the occurrence of clinical symptoms of disease. Conversely, there are advantages in targeted treatment if the drug has side effects, or if there are any worries about the evolution of drug resistant strains of the parasite[52].

We can use our models to calculate the proportional reduction achieved by mass chemotherapy, in quantities such as the average worm burden, the prevalence and the average egg output. We denote the levels of these parameters before control as M^*, P^* and E^* respectively and define the probability that an individual harbours i parasites as $p(i)$, and that he receives treatment as $g(i)$, where

$$g(i) = a[1-(1-\alpha)e^{-i/I}] \qquad (13)$$

a and α (both < 1) are constants defining the upper and lower bounds on $g(i)$, respectively, while I is a measure of the selectivity of the treatment. For example if $I = 40$, ~60% of individuals with 40 worms receive drug treatment. A continuous function for $g(i)$ is more realistic than a step function, as in

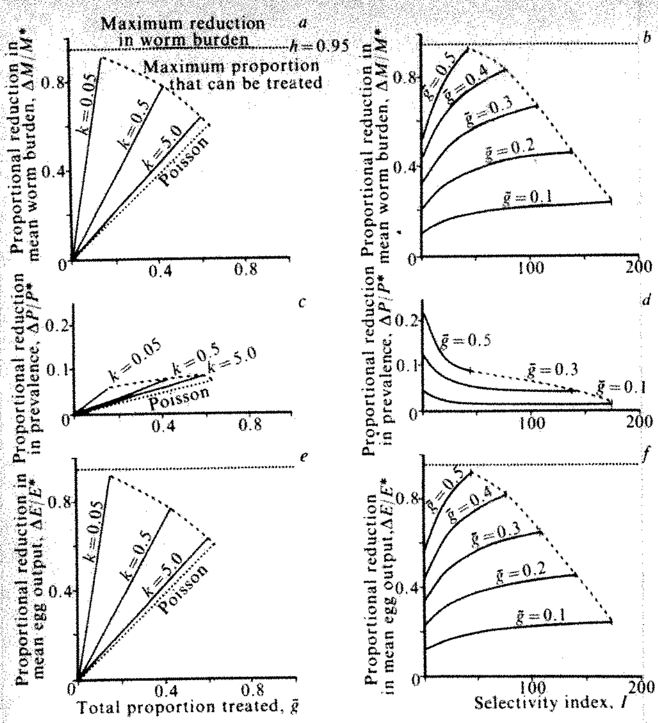


Fig. 5 The impact of selective chemotherapy on the proportional reduction in mean worm burden, $\Delta M/M^*$, the prevalence of infection, $\Delta P/P^*$, and the average egg output per host, $\Delta E/E^*$. *a*, *b*, Relationships between the proportional reduction in average worm burden immediately after a single mass treatment in a community in relation to *a*, the proportion of the population treated, $\bar{g}$, and *b*, the degree of selectivity in treatment, *I*, where the selective function is as defined by equation (13), for various assumptions about the distribution of worms in the human population (that is, various values of the negative binomial clumping parameter, *k*). The relationships in *a* and *b* were calculated using equation (14) with the following parameter values: $M^* = 40$, $I = 40$, $\alpha = 0$, $a = 1$, $h = 0.95$ and $k = 0.5$ (in *b*). *c* and *d*, *f* and *e* show, respectively, similar relationships for the proportional reductions in prevalence and average egg output. The proportional reduction in prevalence, $\Delta P/P^*$ is defined by the relationships $\Delta P/P^* = aJ(I, M^*, k)/[1-(1+(M^*/k))^{k}]$ where $J(I, M^*, k) = [1+(1-h)(M^*/k)]^{-k} - (1-\alpha)[1+(1-y)(M^*/k)]^{-k} - \alpha(1+(M^*/k))^{-k}]$ and $y = h\exp(-1/I)$. The parameter values for *d* and *c* are as defined for *a* and *b*. The proportional reduction in average egg output, $\Delta E/E^*$, is defined by the relationship $\Delta E/E^* = aK(I, M^*, k)/D(I, M^*, k)$ where $K(I, M^*, k) = [[1+(1-x)(M^*/k)]^{-(k+1)} - (1-\alpha)\ y\ [1+(1-xy)(M^*/k)]^{-(k+1)}]$, $D(I, M^*, k) = [1+(1-x)(M^*/k)]^{-(k+1)}$ and $x = \exp(-\bar{b})$ where $\bar{b}$ is as defined in Fig. 2 legend and *y* is as defined above. The parameter values for *e* and *f* are as defined for *a* and *b* excepting *h* is set at unity.

practice there is always considerable uncertainty about the worm burden harboured by any given individual. The proportional reduction in the average worm burden, $\Delta M/M^*$ [where $\Delta M = h\sum_{i=0}^{\infty} ip(i)g(i)$], achieved by treating a proportion $\bar{g}$ [where $\bar{g} = \sum_{i=0}^{\infty} g(i)p(i)$] of the total population with selectivity I is given by

$$\Delta M/M^* = \bar{g}\,h\,\frac{G(I, M^*, k)}{H(I, M^*, k)} \tag{14}$$

where h, M^* and k are as defined earlier. The functions G and H are defined as

$$G(I, M^*, k) = [1-\hat{z}(1-\alpha)[1+(1-\hat{z})(M^*/k)]^{-(k+1)}] \tag{15}$$

and

$$H(I, M^*, k) = [1-(1-\alpha)[1+(1-\hat{z})(M^*/k)]^{-k}] \tag{16}$$

where $\hat{z} = \exp(-1/I)$.

Figure 5 displays the impact of various programmes, using different selectivity levels (I), and various proportions treated ($\bar{g}$) on the average worm burden, the prevalence and the average egg output, given various assumptions concerning the degree of worm clumping in the human population. Three points emerge from this analysis.

(1) Selective treatment is highly beneficial (in terms of the number of treatments administered), provided the worms are highly clumped in their distribution in the host population (in practice this is indeed the case; see Table 2*b*). For example, in the case shown in Fig. 5*a*, where the selectivity index is set at 40 worms per host, a 50% reduction in average worm burden ($M^* = 40$) is achieved by treating 8% of the population when the worms are highly clumped ($k = 0.05$), while 52% of the population must be treated to achieve the same result when the parasites are randomly distributed ($k \to \infty$). (2) There is little benefit to be gained from being too selective (Fig. 5*b*, *d*, *f*). This is fortuitous since it would be difficult in practice to distinguish between individuals who carried say 50 or 75 worms. (3) Of all the three epidemiological variables considered in Fig. 5, selective treatment has the least effect on prevalence. The proportional reduction in prevalence declines as the selectivity index, I, increases (Fig. 5*d*).

The principle conclusion of this analysis is that, despite the extra cost involved in identifying heavily infected individuals, very substantial benefits may be gained from selective treatment, in terms of reducing both the abundance of the parasite (and hence, in general, the frequency of disease symptoms) and the number of treatments administered. This is primarily a consequence of the highly clumped nature of helminth distributions in human communities. The analysis is most applicable to the control of gut-dwelling helminths. For filarial infections, a reduction in worm burden (and hence transmission) may have no short-term impact on disease prevalence as a result of the permanent damage (such as blindness) induced by heavy infections before control. An added problem in the case of tissue-dwelling helminths is that drug-induced parasite mortality may have toxic side effects in those individuals harbouring heavy worm burdens.

Control by vaccination

Although vaccines against helminth infections are not available at present[11,53], we briefly consider the potential benefits to be gained, at the community level, from their development and use. If a safe and effective vaccine is developed, giving protection for an average period of v years, then the proportion of the population, p, that must be immunized per unit of time to reduce R below unity must exceed a critical value $\bar{p}$:

$$\bar{p} = [1-(1/R_0)]v^{-1} \tag{17}$$

For example, if the vaccine gives life protection against a helminth infection with an R_0 value of 3 (appropriate for hookworm in endemic areas, see Table 2), it would be necessary to protect 67% of the community by a single vaccination (or course of vaccination) in infancy. However, if the vaccine only gives protection for a few years, then a large proportion of the population must be repeatedly immunized to sustain the necessary level of herd immunity for community protection. The relative merits of this type of vaccine when compared with cheap antihelmintic agents will therefore depend critically on cost factors since both forms of treatment will have to be administered repeatedly.

Conclusions

A major priority for the community control of infectious diseases is clearly the development of safe, effective and cheap agents (whether drugs or vaccines) to protect the individuals most at risk. These goals have been largely achieved for many of the common infectious diseases and yet they remain endemic throughout large regions of the world[49]. In the less developed countries, the availability of resources for primary health care is a major factor restricting the effective control of these diseases. It would therefore seem desirable to devote effort to elucidating the optimum ways of using the available methods,

for the benefit of the community as a whole. A knowledge of the population dynamics of the disease agent has a central role in this field of research[28,30,33].

Human helminth infections are good examples of the discrepancy between our knowledge of how to treat an individual and how to control the infection in a community. The present analysis suggests that it is possible to determine, in quantitative terms, the level of antihelmintic treatment required in the community either to suppress parasite abundance to a defined level (to eliminate disease symptoms) or to eradicate the infection. More important, given the limitation of resources in areas where such infections are endemic, the analysis indicates that more attention should be given to the practical problems involved in selective treatment because the benefits to be gained by this approach seem very substantial. These benefits will be further enhanced, if, as has been suggested (but as yet unconfirmed[54]), the heavily infected individuals are predisposed to this state, not by the laws of chance, but as a consequence of genetic, behavioural or social factors.

This work was supported in part by US NSF grant DEB81-02783 (R.M.M.) and in part by the UNDP/World Bank WHO Special Programme for Research and Training in Tropical Diseases (R.M.A.). We thank K. Dietz for helpful discussions and comments on the manuscript.

Received 5 January; accepted 18 April 1982.

1. Stoll, N. R. *J. Parasit.* **33**, 1–18 (1947).
2. Iarotski, L. S. & Davis, A. *Bull. Wld Hlth Org.* **58**, 115–127 (1981).
3. Walsh, J. A. & Warren, K. S. *New Engl. Med. J.* **301**, 967–974 (1979).
4. Peters, W. in *The Relevance of Parasitology to Human Welfare Today* (eds Taylor, A. E. & Muller, R.) (Blackwell Scientific, Oxford, 1978).
5. Warren, K. S. *A. Rev. pub. Hlth* **2**, 101–115 (1981).
6. Anderson, R. M. in *Population Dynamics* (eds Anderson, R. M., Turner, B. D. & Taylor, L. R.) 245–282 (Blackwell Scientific, Oxford, 1979).
7. Anderson, R. M. *Lect. Notes Biomath.* **39**, 278–322 (1980).
8. Muller, R. *Worm and Disease: A Manual of Medical Helminthology* (Heinemann, London, 1975).
9. Warren, K. S. & Mahmoud, A. A. F. (eds) *Geographic Medicine for the Practitioner* (University of Chicago Press, 1978).
10. Schad, G. A. & Rozeboom, L. E. *A. Rev. Ecol. Syst.* **7**, 393–420 (1976).
11. Smith, M. A., Clegg, J. A., Snary, D. & Trejdosiewicz, A. J. *Parasitology* **84**, 83–92 (1982).
12. Van den Bossche, H. *Nature* **274**, 626–630 (1978).
13. W. H. O. *Tech. Rep. ser.* No. 615 (1977).
14. Anderson, R. M. & May, R. M. *Nature* **280**, 361–367 (1979).
15. May, R. M. & Anderson, R. M. *Nature* **280**, 455–461 (1979).
16. Croll, N. A., Anderson, R. M., Gyorkos, T. W. & Ghadirian, E. *Trans. R. Soc. trop. Med. Hyg.* **76**, 187–197 (1982).
17. May, R. M. *Math. Biosci.* **35**, 301–343 (1977).
18. Nawalinski, T., Schad, G. A. & Choudhury, A. B. *Am. J. trop. Med. Hyg.* **27**, 1152–1161, 1162–1173 (1978).
19. Wakelin, D. J. *Nature* **273**, 617–620 (1978).
20. Mitchell, G. F. *Adv. Immun.* **28**, 451–511 (1979).
21. Anderson, R. M. & May, R. M. *J. Anim. Ecol.* **47**, 219–248 (1978).
22. May, R. M. & Anderson, R. M. *J. Anim. Ecol.* **47**, 249–267 (1978).
23. Anderson, R. M. *J. theor. Biol.* **82**, 283–311 (1980).
24. Crofton, H. D. *Parasitology* **62**, 179–193 (1971).
25. Macdonald, G. *Trans. R. Soc. trop. Med. Hyg.* **59**, 489–506 (1965).
26. Bradley, D. J. & May, R. M. *Trans. R. Soc. trop. Med. Hyg.* **72**, 262–273 (1978).
27. Fisher, R. A. *The Genetical Theory of Natural Selection* (Clarendon, Oxford, 1930).
28. Dietz, K. in *Population Dynamics of Infectious Diseases: Theory and Applications* (ed. Anderson, R. M.) (Chapman & Hall, London, in the press).
29. Dietz, K. *SIAM-SIMS Conf. Ser.* **21**, 104–121 (1975).
30. Anderson, R. M. & May, R. M. *Science* **215**, 1053–1060 (1982).
31. Anderson, R. M. (ed.) *Population Dynamics of Infectious Diseases: Theory and Applications* (Chapman & Hall, London, in the press).
32. Anderson, R. M. in *Modern Parasitology* (ed. Cox, F. E. G.) (Blackwell Scientific, Oxford, in the press).
33. Dietz, K. in *Population Biology of Infectious Diseases* (eds Anderson, R. M. & May, R. M.) (Springer, Berlin, in the press).
34. Bradley, D. J. & McCullough, F. S. *Trans. R. Soc. trop. Med. Hyg.* **67**, 491–500 (1973).
35. Muench, H. *Catalytic Models in Epidemiology* (Harvard University Press, 1959).
36. Warren, K. S. *J. infect. Dis.* **127**, 595–609 (1973).
37. Smithers, S. R. (ed.). *New Developments with Human and Veterinary Vaccines* 287–299 (Liss, New York, 1980).
38. Wagner, W. H. *Interdisc. Sci. Rev.* **5**, 186–203 (1980).
39. Chavania, A. P., Swalzelder, J. C., Villarejos, V. M. & Zeledon, R. *Am. J. trop. Med. Hyg.* **22**, 592–595 (1973).
40. Willett, W. C., Kilama, W. L. & Kihamia, C. M. *Am. J. Pub. Hlth* **69**, 987–991 (1979).
41. Janckes, M. F., Cornet, P. & Thunpont, D. *Trop. geogr. Med.* **31**, 111–122 (1979).
42. Santos, A. *et al. Bull. Wld Hlth Org.* **57**, 793–799 (1979).
43. Davis, A. & Wiegner, D. H. G. *Bull. Wld Hlth Org.* **57**, 767–771 (1979).
44. Ishizaki, T., Kamo, E. & Boehme, K. *Bull. Wld Hlth Org.* **57**, 787–791 (1979).
45. Davis, A., Biles, J. E. & Ulrich, A. M. *Bull. Wld Hlth Org.* **57**, 773–779 (1979).
46. Warren, K. S. in *Principles and Practice of Infectious Diseases* (eds Mandell, G. L., Douglas, R. G. & Bennett, J. E.) (Wiley, New York, 1979).
47. Arfaa, F. & Ghadirian, E. *Am. J. trop. Med. Hyg.* **26**, 866–871 (1977).
48. Smillie, W. G. *S. med. J., Nashville* **17**, 414–499 (1924).
49. Rougement, A., Boisson, M. E., Borges de Silva, G. & Zander, N. *Bull. Wld Hlth Org.* **54**, 403–410 (1976).
50. Warren, K. S. & Mahmoud, A. A. F. *Trans. Ass. Am. Physns* **89**, 195–204 (1976).
51. Warren, K. S., Siongok, A., Houser, T. K., Ouma, J. H. & Peters, P. A. *J. infect. Dis.* **138**, 849–855; **138**, 856–858 (1978).
52. Araujo, N., Katz, N., Pinto Dias, E. & Pereira de Souza, C. *Am. J. trop. Med.* **29**, 890–894 (1980).
53. Lloyd, S. *Parasitology* **83**, 225–242 (1981).
54. Croll, N. A. & Ghadirian, E. *Trop. geogr. Med.* **33**, 241–284. (1981).
55. Ansari, N. (ed.) *Epidemiology and Control of Schistosomiasis* (Karger, Basel, 1973).
56. Anderson, R. M. & May, R. M. *Parasitology* **79**, 63–94 (1979).
57. Anderson, R. M. in *Vectors of Disease Agents, Interactions with Plants, Animals and Man* (eds McKelvey, J. J., Elridge, B. F. & Maramorosch, K.) 13–43 (Praager, New York, 1981).
58. Krafsur, E. S. & Garrett-Jones, C. *Trans. R. Soc. trop. Med. Hyg.* **71**, 155–160 (1977).
59. Cheever, A. W. *Am. J. trop. Med. Hyg.* **17**, 38–64 (1968).
60. Hesieh, H. C. *Jap. J. Parasit.* **19**, 508–522 (1970).
61. Pesigan, T. P. *et al. Bull. Wld Hlth Org.* **18**, 345–455 (1958).
62. Bhallacharya, N. C. & Gubler, D. J. *Indian J. med. Res.* **61**, 8–11 (1973).
63. Hill, R. B. *Am. J. Hyg.* **6**, 19–41 (1926).
64. Stoll, N. R. *Am. J. Hyg.* **3**, 103–117 (1923).
65. Cheever, A. W., Kaner, I. A., Elwi, A. M., Moismann, J. E. & Danner, R. *Am. J. trop. Med. Hyg.* **26**, 702–716 (1977).
66. Miller, M. J., Krupp, I. M., Little, M. D. & Santos, C. *J. Am. med. Ass.* **230**, 1412–1414 (1974).
67. Wolfe, M. S. & Wershing, J. M. *J. Am. med. Ass.* **230**, 1408–1411 (1974).
68. Cameron, T. W. M. *Parasites and Parasitism* (Methuen, London, 1956).
69. Mahmoud, A. A. F. & Warren, K. S. *Clin. Res.* **28**, 474 (1980).
70. Dietz, K. *Lect. Notes Biomath.* **11**, 1–15 (1976).

Structure of the expanded state of tomato bushy stunt virus

I. K. Robinson & S. C. Harrison

Gibbs Laboratory, Harvard University, Cambridge, Massachusetts 02138, USA

The full three-dimensional structure determination of tomato bushy stunt virus has shown how its protein subunit can accommodate to different packing environments in the viral coat and how the size of the particle is nonetheless determined without ambiguity. Reversible, cooperative expansion of the virus at pH *values above neutrality and in the absence of divalent cation requires further bonding alternatives for the subunit. We report here the structure of the expanded tomato bushy stunt virus particle, determined at 8 Å resolution by X-ray crystallography. Subunit conformational changes are localized, and folded domains are essentially invariant. The bonding properties of the subunit can account for controlled polymorphism of the assembly.*

THE tomato bushy stunt virus (TBSV) particle comprises a single strand of RNA (molecular weight (M_r) 1.5×10^6) and 180 copies of a 40,000 M_r coat protein[1]. The subunits of the coat are arranged in a $T=3$ icosahedral lattice[2] as shown in Fig. 1*a*. High-resolution crystal structure analysis[3] has revealed the positions of all amino-acid residues in the C-terminal 75% of the protein chains; these form a shell having inner and outer radii of 110 and 170 Å. The N-terminal parts of the protein subunits and the whole of the RNA are invisible in the icosahedrally averaged, high-resolution electron density maps; they

0028–0836/82/240563—06$01.00

are thus believed to be spatially disordered. These components together fill the cavity in the interior of the virus[4,5].

The ordered part of the protein folds into two structural domains linked by a short 'hinge'. The 170-amino acid 'S'-domains pack tightly into the shell between radii of 110 and 140 Å. The three symmetrically distinct positions of this domain, shown in Fig. 1*b*, are denoted A, B and C. They have identical folding of their polypeptide backbone, but adjust to three 'quasi-equivalent' environments by differences in conformation of certain side chains on their surfaces and, more dramatically, by the ordered folding of 35 residues at the N-terminal end of the 60 C-position S-domains. These 35 residues form the so-called 'extended arm' and 'β-annulus'[3]. The A- and B-position S-domains make an extensive contact with each other, which is broken in C positions by interposition of the arm. Distal residues of the arm interdigitate with corresponding residues of two other C subunits to form the β-annuli. A coherent internal scaffold is thereby created, which precisely determines the curvature and size of the virus particle.

The second domain, called 'P', forms strong dimers that lie on the 2-fold and quasi-2-fold axes of the constellation of S-domains. These dimers project to the outer radius of 170 Å and are clearly visible in electron micrographs of the virus[6].

Expansion

TBSV and many other related viruses undergo a structural phase transition in solution that is controlled primarily by *p*H and divalent cation concentration. The phenomenon was first observed in bromegrass mosaic virus (BMV) by Incardona and Kaesberg[7] using various physicochemical techniques. These all gave results consistent with a model in which the particle mass was conserved and the hydrodynamic radius increased by ~10%. For this reason we will refer to it as 'expansion' of the virus.

BMV exhibits an anomalous titration of two protons per subunit associated with the phase transition at *p*H 6.7. This has been interpreted[8,9] as due to formation of hydrogen bonds

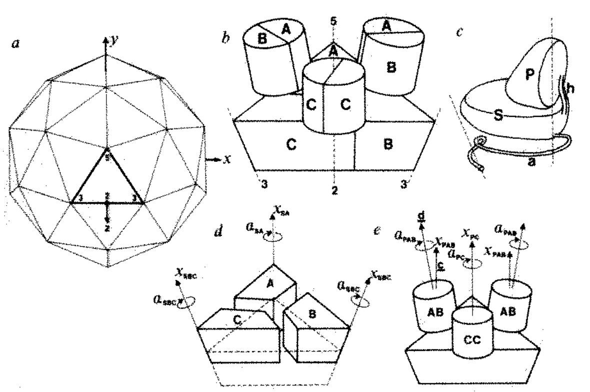


Fig. 1 Architecture of the TBSV particle as determined by high-resolution X-ray structure analysis, and model of its expansion. *a*, Stellated pentagonal dodecahedron showing the definition of the icosahedral asymmetric unit (highlighted) as the spherical sector lying between a 5-fold and two adjacent 3-fold axes. The viral shell consists of 60 of these identical units. *b*, Schematic diagram showing the packing of three quasi-equivalent TBSV protein subunits in the icosahedral asymmetric unit, denoted A, B and C. *c*, Approximate shape of the C-position subunit, showing the two domains, denoted 'S' (surface) and 'P' (projecting), connected by a polypeptide hinge, h. The N-terminal arm, a, is ordered only in the C-position subunit and disordered in A and B. *d*, Parameterization of the model of the expansion of the S-domains. A translation, *x*, parallel to the symmetry axis and a rotation, α, about that axis is the most general description. The B- and C-position domains are assumed to retain their contacts about the 3-fold axis and so are described by the same parameter pair. *e*, Parameterization of the P-domain expansion. The AB P-domain dimer is unconstrained by the particle symmetry, and so has the full six parameters of a rigid body transformation. Spherical polar coordinates are used for compatibility with the other parameter definitions.

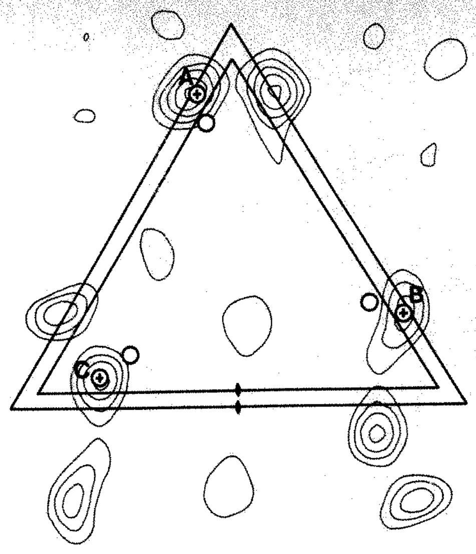


Fig. 2 A section of the 12 Å difference Fourier for the $PtCl_4$ derivative, located 132 Å from the particle centre. Phases for this map were derived from the best three-parameter model and refined by symmetry averaging. The larger triangle denotes the boundary of the icosahedral asymmetric unit and the symbol ⊕ denotes the refined atomic positions. The quasi-symmetry of these site positions is clearly apparent. Superimposed on the same scale are the refined platinum positions for the compact virus (○), which lie at a radius of ~116 Å, and the size of the icosahedral asymmetric unit at that radius (smaller triangle). The vector joining the platinum position in the compact virus to that of the expanded virus is an accurate measure of the displacement of the subunit to which the atom is bound.

between apposed carboxylic acid groups. Divalent cations, known to inhibit the expansion of BMV, would displace the hydrogen bonds in favour of ionic links stable at all *p*H values. Native preparations of TBSV have two bound, EDTA-chelatable ions per subunit (probably both calcium) which must be removed before expansion can be observed[10]. Once free of these ions, the virus expands reversibly when the *p*H is raised above 7.0; the transition seems to be characterized by some hysteresis, which is most easily followed by changes in the solution X-ray scattering. The location and structure of the cation binding sites[11] are consistent with the mechanism just described: five aspartate side chains, contributed by each pair of adjacent S-domains in the trimer of Fig. 1*b*, are involved in the binding of the two ions. When charged at high *p*H, these aspartates repel each other and destabilize the inter-domain contact. This interface is therefore likely to be the one broken to permit expansion.

Examination of the relative extent of various contacts between S-domains in the high-resolution structure[3,11] led to a model for expansion, shown in Fig. 1*d*. This model assumed that the virus retained its icosahedral symmetry and that the internal tertiary structure of each domain was conserved. Rapid reversibility of expansion further suggested that the interlocked internal scaffold was conserved and hence that the pentamer and hexmer clusterings of S-domains in the virus were unaltered. As argued above, it was reasonable to suppose that the trimer contacts, which contain the divalent cation sites, were disrupted. With this model, S-domain displacements can be described by just four parameters, two rotations about and two translations along the symmetry axes. The motions of the P-domains described in Fig. 1*e* are rather more general: only the PP dimer contact is conserved, so eight parameters make a complete description. We have used this model-building approach in the crystal structure determination.

Structure determination

Crystals of the expanded virus were prepared by vapour diffusion[12] from 2.9% polyethylene glycol, 0.01 M EDTA, 0.058 M $(NH_4)_2HPO_4$, 20% ethylene glycol, *p*H 7.5 at 18 °C.

X-ray diffraction showed crystalline order to ~7 Å resolution. The space group is C2, with cell constants $a = 546.3$ Å, $b = 433.1$ Å, $c = 383.4$ Å and $\beta = 134.0°$. The crystallographic asymmetric unit contains one half of a virus particle, so there is 30-fold non-crystallographic symmetry, which can be exploited for refinement of phases[13]. The particle position within the unit cell is defined in this space group with the exception of the particle orientation angle about the 2-fold axis, referred to as θ, which must be determined experimentally; an additional parameter must therefore be added to the model description of the expanded virus in the crystals.

Data were collected by oscillation photography[14] for native crystals as well as $PtCl_4^{2-}$ and ethyl mercury thiosalicylate (EMTS) heavy-atom derivatives. These heavy-atom reagents were similar to those used in the structure determination of the compact virus[15], in the hope that the same sites would be labelled and a direct comparison of coordinates could be made. The data quality deteriorated sharply with resolution, and only reflections in the range 8–35 Å were considered usable. Table 1*a* summarizes native and derivative data collection. Comparison of Wilson plots for data from the compact and expanded forms of the virus indicated that the absence of high-resolution data for the latter was well modelled by gaussian disorder, with a Debye 'temperature factor' of 340 Å^2.

Starting values of the model parameters were obtained by analysing the packing of the expanded virus particles into the crystal unit cell. These were used to compute a predicted diffraction pattern, which was compared with the observed pattern by calculating an R-factor for agreement of scaled intensities. The model parameters were adjusted until the R-factor was minimized. A value of $R = 0.52$ was obtained for 12 Å, significantly lower than the value of 0.59 obtained by comparing sets of uncorrelated intensities. A more detailed account of this critical step and of the rest of the structure determination will be published elsewhere.

Phases calculated from the best model were applied to the observed structure facors to 12 Å resolution and subjected to refinement by non-crystallographic symmetry averaging[13]. The R-factor for comparison of symmetry-averaged and observed intensities was reduced to 0.38 in six cycles (12 Å data) and the phases on average moved by 55°.

To confirm the validity of the phasing and to extend its resolution, heavy-atom difference maps were calculated for both derivatives using the refined 12 Å phases. An example is shown in Fig. 2. Sites corresponding to the three platinum positions[15] were unambiguously identified. Their parameters were refined by conventional methods[16]. The *a priori* assumptions of rigid-body domain motion and conservation of hexamer and pentamer subunit contacts were supported by the finding that the positions of the S-domain heavy-atom sites were self-consistent to within 2 Å (Table 1*b*). The refined positions allowed a considerably more precise set of model parameters to be derived directly.

The three steps of model-building, R-factor minimization and non-crystallographic symmetry phase refinement were then repeated to extend the phasing to the full 8 Å data set. The refined heavy-atom occupancies at 8 Å were unchanged from their values at 12 Å (Table 1*b*), indicating the absence of significant deviations from icosahedral symmetry. Fourier analysis of the best model gave an R-factor of 0.46 with all the observed data. Phase refinement brought this value down to 0.43; the lack of dramatic improvement here indicates that the model description is a good one and reflects the poorer quality of the higher-resolution data. Electron density maps were prepared with dots superimposed for every amino acid residue of the model structure. Corresponding 8 Å maps of the compact structure with an imposed 340 Å^2 'temperature factor' were also prepared for comparison.

Interpretation of electron density

The successful refinement and correct determination of heavy-atom positions confirm our assumption that tertiary structure in each domain is conserved, with conformational changes restricted to localized flexible regions. The displacements and rotations generated by these local flexions are described by the refined model parameters in Table 2. A striking aspect of the expanded state is the appearance of a branched opening 80 Å long and large enough for a 20 Å sphere to pass through each of the 60 faces; this is seen in Fig. 3, and an overall view of the icosahedral asymmetric unit is shown in Fig. 4*a*. The P-domains are rotated by relatively large angles, 30° in the AB case and 103° for the CC dimer. This latter value is so great that the actual direction of the rotation is ambiguous; the clockwise direction is chosen on the basis of a more reasonable conformation of the hinge. To accommodate these rotations and to bridge the gap between the domains, it is necessary to postulate a new 'expansion hinge', several residues long, to the N-terminal side of the hinge associated with the quasi-symmetry[3]. We have thus seen four distinct states of the linkage

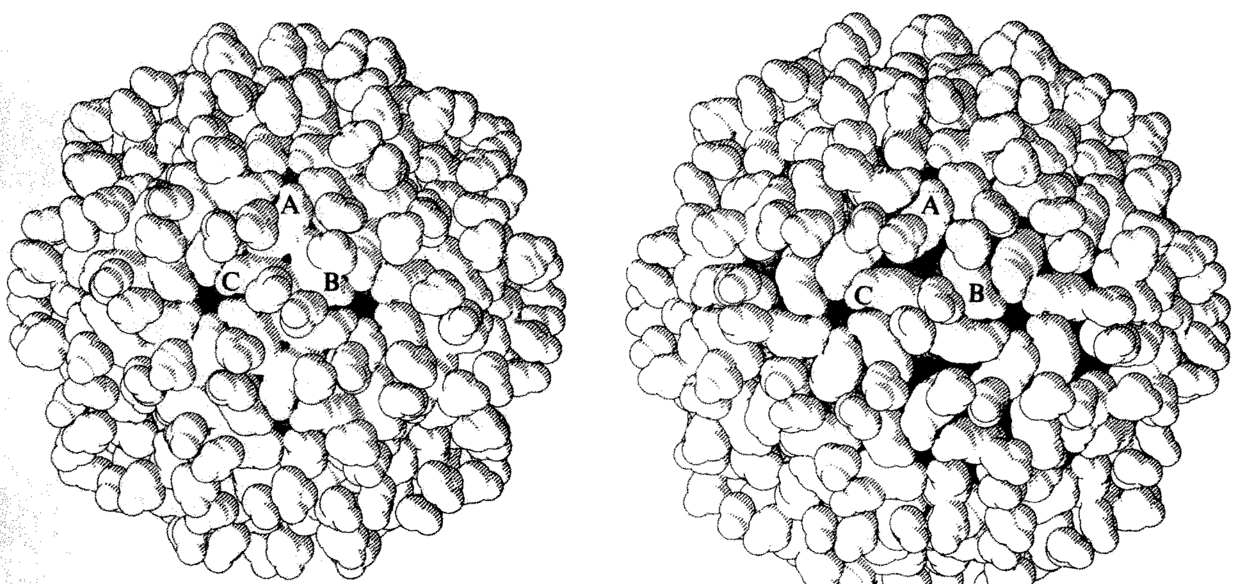


Fig. 3 Pictorial view of the expansion of TBSV. The compact particle is on the left. Note the appearance of the inter-subunit opening and the rotation of the projecting domains in the expanded virus on the right. (This figure was prepared by A. J. Olson using a modified form of the PLUTO program of S. Motherwell.)

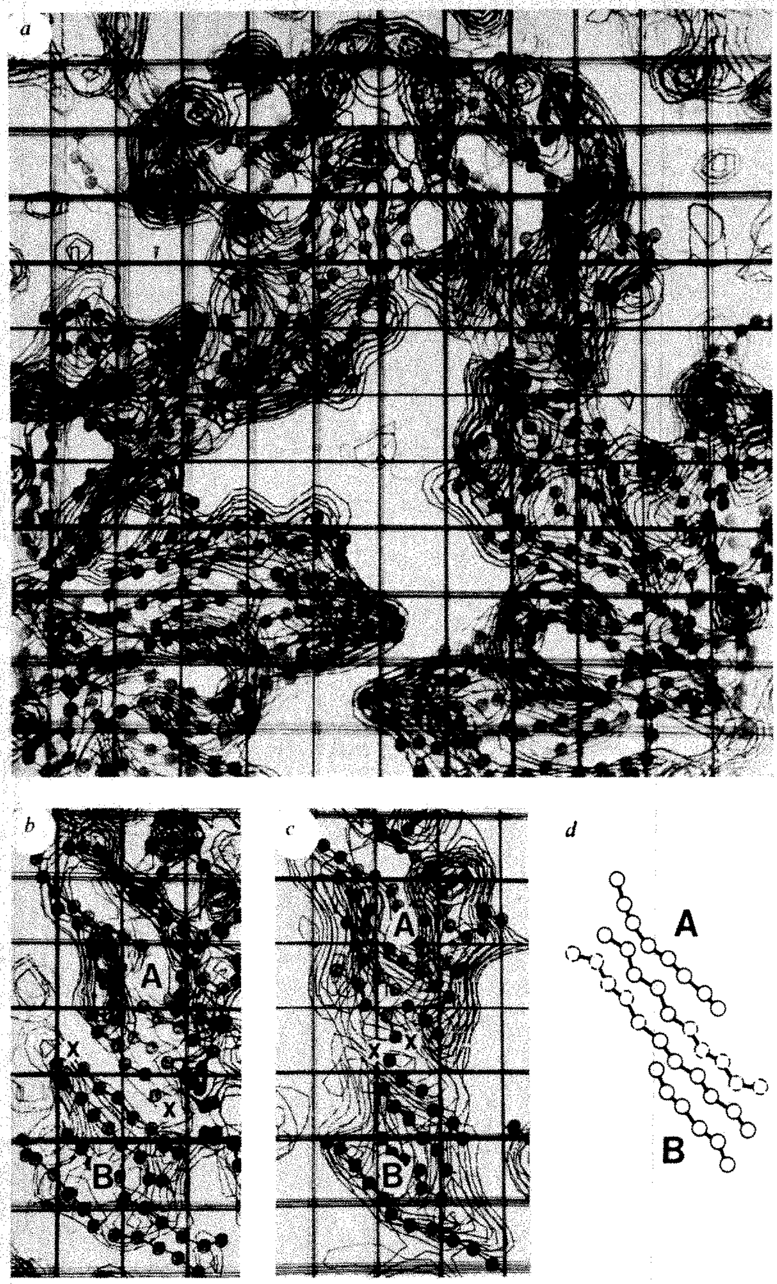


Fig. 4 Electron density maps at 8 Å resolution and their interpretation. Dots are superimposed for each amino acid in the best-fitting model of the expanded virus. The grid is 10 Å square. *a*, Overall view of the icosahedral asymmetric unit of expanded TBSV in the S-domain region. The opening is clearly visible between the domains. An impression of the quality of the map can be obtained from the fit of model to density and from the noise level in the solvent regions. *b*, The contact between the A and B position S-domains of the compact structure. The residue marked 'X' in each subunit is the first residue to be included at the N-terminus of the model. This part of the structure contains a four-strand antiparallel β-sheet extending across the base of the domain. *c*, The corresponding region of the expanded structure in which the two domains are shifted past each other by 14 Å. There is clear density immediately preceding residue X which can be interpreted as an ordering of the arm. *d*, Interpretation of the new electron density as an additional strand of β-sheet that forms a clamp between the subunits.

region between S- and P-domains: in A/B and in C positions in each of the compact and expanded particles.

The refined expansion parameters and the known high-resolution structures of the individual domains define an accurate model of the expanded virus. Use of this model greatly facilitated further interpretation of the refined electron density. Even though the resolution was not sufficient to distinguish individual strands of protein, the fit of model to density was good enough for small changes from the model to be detected reliably.

The extended arm, which is ordered in C-position S-domains but not in A or B, is conserved in the expanded virus, as is the β-annulus. The latter feature is much stronger in the expanded form; it appears ordered for several residues further into the interior of the virus. The function of the C-position arm and the β-annulus is to determine the size of the virus during assembly[3,11]. Conservation of these features in the expanded state is consistent with their fundamental role in establishing structural coherence.

The contact between the four-stranded β-sheets at the base of the A- and B-position S-domains is sheared apart by the expansion, as shown in Fig. 4*b* and *c*. A very distinct bulge of new density lies beside the first strand in both the A and B positions of the expanded state. This seems to be additional ordered protein at the N-terminus of the domain, and we suggest the interpretation of Fig. 4*d*. About six residues immediately N-terminal to the A and B position, S-domains of the compact structure become ordered in the expanded state, extending the first strand of each β-sheet and forming hydrogen bonds with residues of the opposite subunit. This interaction would create a single large sheet extending across the inner surface of both S-domains (Fig. 5). The density is interpreted in this way as the new B-position arm extending beyond the base of the A-position S-domain for a further 20 Å along the inside edge of the inter-subunit opening, terminating at a radius of 156 Å, close to the exterior of the virus. Thus, one of the arms seems to pass all the way through the opening to the outside, possibly explaining the observation that mild proteolysis of expanded virus causes a cleavage of many of the chains at positions 70 to 100 residues from the N-terminus[17].

A feature of the S-domain also seems to have shifted during expansion. The trimer contact region in the centre of the triangular icosahedral asymmetric unit of the virus (see Fig. 1*b*) consists of three loops of extended chain in symmetrical contact. In the maps of expanded virus, the rigid model of the domain has no electron density in this region, while a new loop of density of equal contour length protrudes into the opening. We suggest that a local refolding of the chain occurs in this region, with a displacement of ~10 Å at the tip of the loop. A similar

change in secondary structure of an exterior loop has been observed in trypsinogen[18].

Apart from this limited number of modifications, which are easily rationalized, the model of the expanded virus is completely consistent with the electron density. As all the density is accounted for, we are confident of the interpretation.

Stability and disorder in the expanded structure

The detailed interpretation of electron density described above shows that the expanded state is characterized by a relatively rigid network of S-domains. This network may be stabilized in part by a new contact between the bases of the A and B S-domains, involving an ordered conformation for about six residues of the arm in those positions where it is not ordered in the compact state. The extended arm and β-annulus of the C-position subunits are conserved. The P-domains are rearranged on the surface of the virus by fairly large rotations about their local diad axes.

Table 1 Data collection statistics for expanded TBSV (a) and refined heavy-atom coordinates (b)

a

	16	12	8	
Resolution limit (Å)	16	12	8	
No. of possible reflections	6,888	17,323	60,193	
No. observed	3,880	10,356	36,113	Native
R_{sym}	0.10	0.12	0.20	Native
No. observed	508	1,357	4,733	$PtCl_4$
R_{sym}	0.10	0.12	0.16	$PtCl_4$
No. observed	1,351	3,607	12,579	EMTS
R_{sym}	—	0.11	0.20	EMTS

b

	x	y	z	Occupancy (12 Å)	Occupancy (8 Å)
Platinum					
A	−8	67	134	8.1	8.0
B	37	19	133	6.4	7.2
C	−30	7	133	6.7	7.0
Mercury					
A1	−17	44	146	4.1	4.2
B1	19	40	145	4.5	2.9
Cl	−5	9	145	3.2	2.7
A2	−29	48	178	2.2	4.5
B2	29	46	179	3.8	4.6
C2	10	5	179	2.1	2.2
E	−43	1	151	3.2	3.1

a, A total of 56 1° oscillation films were used in the final native data set. The reproducibility of the data is indicated for the redundant measurements by R_{sym}.

$$R_{sym}=\frac{\sum_h \sum_{j=1}^{n_h} |I_{hj}-\bar{I}_h|}{\sum_h n_h \bar{I}_h}$$

where the I_{hj}, $j=1,\ldots n_h$ are the repeated measurements of I_h, and $\bar{I}_h$ is their mean. *b*, Refined heavy-atom coordinates (Å in coordinate frame of Fig. 1*a*) and relative occupancies (arbitrary units). Individual heavy-atom temperature factors were set to 25 $Å^2$ and not refined. Occupancies are shown for refinements at 12 Å and 8 Å resolution. All sites except E have chemically equivalent positions in the derivatives of compact TBSV, and their nomenclature corresponds to that of Table 2 in ref. 15. Site E is unique to the expanded structure and lies near a known cysteine.

Table 2 Final parameters for the best model of expanded TBSV

Translations		Rotations	
x_{SA}	19.4 Å	α_{SA}	4.5°
x_{SBC}	17.8 Å	α_{SBC}	2.1°
x_{PAB}	18.1 Å	α_{PAB}	30°
x_{PC}	14.2 Å	α_{PC}	103°

The parameters are defined in Fig. 1*d* and *e* legends. Direction cosines for AB P-domain dimer translation and rotation. $\vec{c}$, (−0.187, 0.352, 0.917); $\vec{d}$, (−0.264, 0.140, 0.954); θ, 97.1°.

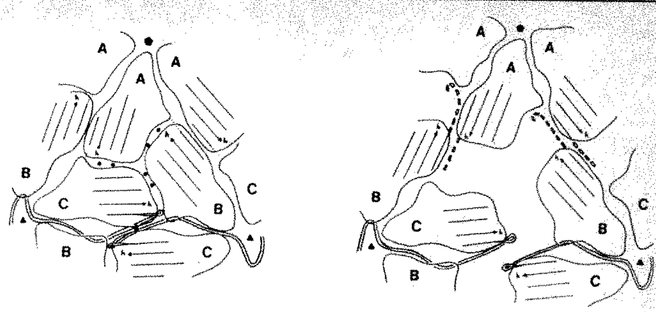


Fig. 5 Schematic diagrams showing the inner surface of S-domains in compact and expanded TBSV and indicating the interpreted disposition of the A and B position arms in the expanded structure. The Ca^{2+} sites are shown by pairs of dots[10]. The β-sheets at the base of the S-domains are denoted by four parallel lines, the C arms by double lines and the interpretation of A and B arms in expanded TBSV by broken lines. In this interpretation, the A arm extends to make a six-residue contact with the B subunit, and the B arm extends for about 15 residues, making the same 6-residue contact with the A subunit and continuing up the inside of the opening to the outside of the virus. The C arm has exactly the same conformation in both structures, except that in expanded virus it is ordered ~10 Å further into the viral interior. The strands terminating in arrows and indicated by 'h' lead to the inter-domain hinge. The change in spacing of these arrowheads across dyads when the particle expands illustrates the requirement for an 'expansion hinge' as described in the text.

The CC P-domain dimer appears to be more disordered than the rest of the structure: the corresponding heavy-atom site refined rather poorly, and the resulting electron density of this domain is considerably weaker than that of the quasi-symmetry-related AB dimer. A model was built with individual relative isotropic temperature factors for the various domains and optimized with respect to the observed dta. The best model had $B=80$ $Å^2$ for the A and B position P-domains and $B=600$ $Å^2$ for the C position. The latter temperature factor corresponds to a r.m.s. positional fluctuation relative to the S-domains of 5 Å (ref. 19), and because this domain is involved in the inter-particle contact of the crystals it can explain the disorder of the crystals as a whole.

Because of the disorder of the P-domains, it seems unlikely that they are involved in a rigid way with clamping the particle in the expanded state. Moreover, southern bean mosaic virus (SBMV), which has a single domain structurally homologous to the TBSV S-domain and no P-domain[20], also has an expanded particle[21]. If the structural similarity of these viruses extends to the expanded state, the P-domain cannot be specifically involved in the expansion mechanism, and its motions must be a consequence of more fundamental structural changes occurring elsewhere in the virus. The proposed AB S-domain contact (Fig. 5) is a feature that could equally well exist in the expanded state of SBMV as in TBSV.

There is no evidence for ordering of the RNA or the N-terminal portion of the protein, other than the places mentioned. The inter-subunit opening is sufficiently large for the arm to be able to extrude entirely from the virus, even if the N-terminal 65 residues were folded as a third domain of the structure. That some arms do extrude is consistent with their proteolytic sensitivity in expanded TBSV[17].

The expansion mechanism is triggered by deprotonation of the aspartate residues of the calcium binding sites, provided the Ca^{2+} salt bridges have been removed by chelation (Fig. 5). The local build-up of charges must then prise the opening apart. The repulsive force, which would increase monotonically with pH in the absence of any induced structural change, is counterbalanced by the attractive forces between domains and between the RNA and the basic amino acids of the interior surface. Beyond a critical pH, the balance swings in favour of expansion, and the particle switches to its other state by rearrangement of

the S-domains. The P-domains then presumably follow suit by adopting minimum-energy configurations on the new surface.

Functions of the structural change

The physiological significance of expansion, if any, is unknown. The expanded state, unlike the compact one, is very sensitive to enzymatic proteolysis, just as it is for BMV[22]. In both cases, this seems to be due to accessibility of normally buried N-terminal sequences. These common features may reflect a common route of entry and disassembly, and Durham[23] has suggested that a difference in calcium concentration between the cytoplasm and the extracellular fluid might be significant in this regard. It is also possible that the expanded state is an intermediate in the virus self-assembly. Whatever its role, the expanded particle demonstrates three additional conformations of the viral subunit. As in the compact structure, the polymorphism is restricted to specific regions of the polypeptide that form specific links between globular domains.

The expansion of TBSV is a highly cooperative change from one state of the viral shell to another: indeed, it has the character of a first-order phase transition[17,24]. There is restructuring of the 2-fold contacts, with conservation of the 5- and 6-fold contacts, giving a two-state character to the transformation. Observed deviations from formal two-state behaviour[17,24] are probably due to changes in the spatially disordered part of the structure, particularly in the N-terminal area. We have strong evidence from the cited proteolytic cleavage experiments[17] that some of these arms extend outside the expanded particle in certain conditions. The manner in which restructured non-covalent contacts are built into the subunit design suggests that the expanded form has a function in the virus life cycle. Indeed, expansion seems to be a general property of small RNA plant viruses. There might also be some formal analogy with the reversiblly interconvertible isoelectric forms of small RNA animal viruses[25].

We thank J. M. Hogle, A. J. Olson, C. S. Steele and D. C. Wiley for advice, discussion and assistance. This research was supported by NIH grant CA-13202.

Received 14 October 1981; accepted 30 March 1982.

1. Ziegler, A., Harrison, S. C. and Leberman, R. *Virology* **59**, 509 (1972).
2. Caspar, D. L. D. & Klug, A. Cold Spring Harb. Symp. quant. Biol. **27**, 1 (1962).
3. Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K. & Bricogne, G. *Nature* **276**, 368 (1978).
4. Chauvin, C., Witz, J. & Jacrot, B. *J. molec. Biol.* **124**, 641 (1978).
5. Munowitz, M. G., Dobson, C. M., Griffin, R. G. & Harrison, S. C. *J. molec. Biol.* **141**, 327 (1980).
6. Finch, J. T., Klug, A. & Leberman, R. *J. molec. Biol.* **50**, 215 (1970).
7. Incardona, N. L. & Kaesberg, P. *Biophys. J.* **4**, 11 (1964).
8. Bancroft, J. B. *Adv. Virus Res.* **16**, 99 (1970).
9. Caspar, D. L. D. *Adv. Protein Chem.* **18**, 37 (1963).
10. Hogle, J. M., Kirchhausen, T. & Harrison, S. C. *J. molec. Biol.* (submitted).
11. Harrison, S. C. *Biophys. J.* **32**, 139 (1980).
12. McPherson, A. *Meth. Biochem. Analysis* **23**, 249 (1975).
13. Bricogne, G. *Acta crystallogr.* **A30**, 395 (1974).
14. Arndt, U. W. & Wonacott, A. J. *The Rotation Method in Crystallography* (North-Holland, Amsterdam, 1977).
15. Winkler, F. K., Schutt, C. E., Harrison, S. C. & Bricogne, G. *Nature* **265**, 509 (1977).
16. Dickerson, R. E., Kendrew, J. C. & Strandberg, B. E. *Acta crystallogr.* **14**, 1188 (1961).
17. Sorger, P. K., Stockley, P. G., Robinson, I. K. & Harrison, S. C. (in preparation).
18. Stroud, R. M., Kossiakoff, A. A. & Chambers, J. L. *A. Rev. Biophys. Bioengng* **6**, 177 (1977).
19. Debye, P. *Ann. Phys.* **43**, 49 (1914).
20. Abad-Zapatero, C. *et al. Nature* **286**, 33 (1980).
21. Hsu, C. H., White, J. A. & Sehgal, O. P. *Virology* **81**, 471 (1977).
22. Pfeiffer, P. & Hirth, L. *FEBS Lett.* **56**, 144 (1975).
23. Durham, A. C. H., Hendry, D. A. & von Wechmar, M. B. *Virology* **77**, 524 (1977).
24. Kruse, J. & Witz, J. (in preparation).
25. Rueckert, R. R. *Comp. Virol.* **6**, 131 (1976).

LETTERS TO NATURE

Discovery of 69 ms periodic X-ray pulsations in A0538–66

G. K. Skinner*, D. K. Bedford*, R. F. Elsner†, D. Leahy†, M. C. Weisskopf† & J. Grindlay‡

* Department of Space Research, University of Birmingham, Birmingham BI5 2TT, UK
† Space Sciences Laboratory, NASA Marshall Space Flight Center, Alabama 35812, USA
‡ Center for Astrophysics, 60 Garden St, Cambridge, Massachusetts 02138, USA

Both the recurrent X-ray transient A0538–66 and its optical counterpart undergo outbursts at intervals which are multiples of 16.6 days[1–3]. The times of outbursts are subject to appreciable jitter and although typically brief (<1 day) the event can, on occasions, last nearly a complete cycle[4]. The system lies in the direction of the Large Magellanic Cloud (LMC) and the Doppler shifts of the quiescent-state optical absorption lines indicate that it is a member of the LMC[5,6]. This implies an X-ray luminosity during outburst of 8×10^{38} erg s^{-1} (2–17 keV)[1]—higher than that of any other 'stellar' X-ray source. We report here the detection, during observations with the Einstein Observatory at the time of an outburst, of X-ray pulsations with a period of 69.2126 ms from A0538–66, adding to the unique properties of this system that of being the fastest X-ray pulsar known to be in a binary system. The observed rate of change of period shows the source to be an eccentric binary system in which the two components approach close to each other during the outbursts and may even become immersed in a common envelope.

Two Einstein observations of A0538–66 were made as part of the Guest Observer Program. One of these was made on 16 December 1980 23.73 to 24.85 h UT, at phase 0.056–0.058, where phase zero corresponds to the predicted time of a peak[7]. A point source coincident with the position of A0538–66 was detected by the High Resolution Imager (HRI) in the focal plane of the telescope at a mean level of (0.090 ± 0.005) HRI counts s^{-1}. Unlike the HRI, the Monitor Proportional Counter (MPC) which is aligned with the main telescope and has a field of view of 42 arc min (FWHM) is sensitive to relatively high-energy X rays (1.1–21 keV) and provides energy resolution. The MPC registered a flux ~40 MPC counts s^{-1} above background. The spectrum of the source detected by the MPC showed a strong low energy cutoff corresponding to photoelectric absorption in a column density $\cong 2\times10^{23}$ cm^{-2} (using Brown and Gould[8] cross-sections) making this flux consistent with that observed from A0538–66 with the HRI. Simultaneous fluctuations on time scales 10–1,000 s in the two flux levels, apparently due to variations in the column density of absorbing matter, confirm that the MPC flux was from A0538–66.

The MPC includes a time interval processor (TIP) which records the time between events in the 1.1–21 keV energy band with a resolution of 1 µs or 1.6%, whichever is the greater. The TIP data were analysed by converting to photon arrival times, correcting for the motion of the satellite and of the Earth, and examining them for manifestations of variability. This procedure revealed the existence of a persistent periodicity, detected throughout the entire observation at a mean heliocentric pulse period of 0.0692126 s. Epoch folding the data at the best period and correcting for the existence of a period derivative leads to a result 111σ above the noise, while folding at neighbouring but statistically independent periods produces results consistent with noise.

The pulse period was determined by means of pulse arrival time analysis. Initially the entire data set was folded modulo different trial periods and the χ^2 statistic was maximized as a function of period. The resulting light curve was used as a template and each 208 s section of the data was compared with it using a cross-correlation technique, allowing a relative pulse arrival time, t_m, to be determined. The t_m values were fitted by functions of the form $t_m = t_0 + mP$ and $t_m = t_0 + mP + \frac{1}{2}m^2P\dot{P}$. The fit assuming a constant period was unsatisfactory ($\chi^2 = 52$

for 16 d.f.). Figure 1*b* shows the residuals for this fit and also the curve corresponding to the fit obtained once a period derivative was included, which was excellent ($\chi^2 = 17.7$ for 15 d.f.). The parameters for the best fit obtained are: pulse period $P = (0.06921166 \pm 0.00000017)$ s, at $t_0 =$ JD 2444590.4792, rate of change of period $\dot{P} = (5.01 \pm 0.86) \times 10^{-10}$. All times are heliocentric. Errors are $\pm 1\sigma$.

Figure 1*a* shows the light curve obtained by folding the data on the basis of the best fit P, $\dot{P}$. After allowing for the background, the pulsed component accounts for 26% of the flux from the source. The amplitude of the pulsations in the separate 208s intervals follows the mean count rate above background so that there is no evidence for the pulsed-fraction changing; the pulsed flux is apparently affected by the variable absorption column density in the same way as the steady flux.

The second scheduled Einstein observation of A0538 − 66 took place on 3 February 1981 23.78–25.76 h UT at phase −0.002 to +0.004 (no data were available for 24.28–25.00 h UT). In both the HRI and MPC data A0538 − 66 was seen to be turning on during the last ~1,400 s, averaging 10 MPC counts s^{-1} during this time. No pulsations were detected although they would have been if the pulse fraction and shape had been the same as during the more intense first observation. However, the energy spectrum was considerably harder. The photon power law index was −0.6 during the second observation, compared with −1.6 during the first, although the average column densities were similar. We note that variations of pulse fraction and shape with intensity were found in Uhuru observations of Cen X-3 (ref. 9).

A0538 − 66 was within the field-of-view of the Einstein telescope on two other occasions: 10 April 1979 12.73–13.70 h UT (phase −0.002 to +0.002) and 30 May 1980 9.07–10.63 h UT (phase +0.024 to +0.026). Despite the fortuitous proximity to phase zero, in neither case was significant flux detected with either the MPC or the main telescope. This may be related to the apparent absence of any large scale activity during the optical observations of Pakull and Pamar[5] in 1980.

The relationship $P \gtrsim (G\bar{\rho})^{-1/2}$ between pulsation period and the mean density of the region in which it arises shows that the seat of the 69 ms pulsations must involve a neutron star (or a black hole). The observed $\dot{P}$ is smaller by a factor of $>10^3$ than

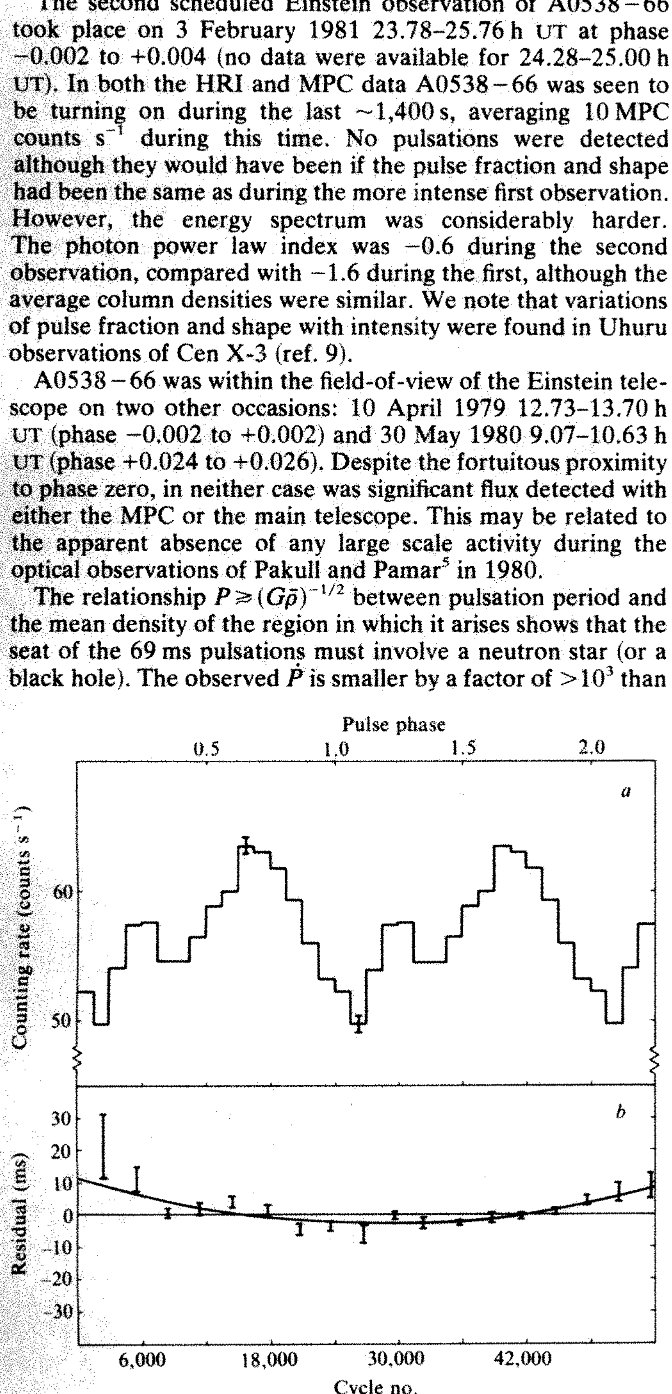


Fig. 1 *a*, The MPC data folded according to the best fit parameters. *b*, Measured pulse arrival time minus that expected on the assumption of a constant period $P = 0.06921264$ s. The curve corresponds to $P = (0.06921166 \pm 0.00000017$ s; $\dot{P} = 5.01 \pm 0.86 \times 10^{-10}$. Error bars are $\pm 1\sigma$.

that of the oscillations reported by Sadeh *et al.*[10] during a burst from MXB1728-34 and we believe that the system is a true pulsar in which the clock is a rotating neutron star.

If the observed $\dot{P}$ were to correspond to a changing rate of rotation, a power of ~10^{41} erg s^{-1} would have to be dissipated in the system. As this seems unreasonable, and the expected time scale for spin-up or spin-down due to accretion torques is too long (see below), we conclude that the major component of the $\dot{P}$ term is due to a changing Doppler shift. Consequently, the neutron star must be in a gravitational field whose line of sight component causes an acceleration of 220 cm s^{-2} and, as optical evidence suggests that the primary has mass $M_{\rm opt} = 12\, M_\odot$, the stellar separation can be no more than 39 $R_\odot$. Charles *et al.*[11] have shown that the region responsible for the optical flux during outburst must be at least 40 $R_\odot$ in radius and so, whether the optical emission is centred on the primary or the secondary, it is already clear that during the outburst we are dealing with two objects essentially in contact.

The maximum acceleration of the secondary in a binary system with period $P_{\rm orb}$ and eccentricity e is

$$(2\pi)^{4/3} G^{1/3} (1-e)^{-2} M_{\rm opt}^{1/3} (1+q)^{-2/3} P_{\rm orb}^{-4/3}$$

where q is the mass ratio $M_{\rm X}/M_{\rm opt}$. If we assume that the orbital period is 16.6 days, then we find $e \geq 0.4$ (we have taken $M_{\rm opt} = 12 M_\odot$, $q = 0.1$, although the conclusion depends only weakly on the choice). The limit $e = 0.4$ corresponds to the case where the observation was made exactly at periastron and the inclination and orientation of the orbit are such that the neutron star was directly between the observer and the primary. Any other configuration implies an even higher eccentricity.

While not losing sight of the possibility that the source of energy might be rotational, through a pulsar mechanism, we now consider the implications of powering the X-ray source by accretion onto the neutron star. The first question is whether accretion is possible in the presence of a magnetic field rotating with such a high angular velocity. Before accreting onto the neutron star, matter must pass through a co-rotating magnetosphere. If the radius of the magnetosphere is larger than that of the keplerian orbit co-rotating with the neutron star, accretion is unlikely. The size of the magnetosphere will depend not only on the magnetic field but also on the amount of inflowing matter and so on the X-ray luminosity. We deduce

$$B \leq f \times 10^{11} \left(\frac{L_{\rm X}}{10^{39}\ {\rm erg\ s^{-1}}}\right)^{1/2} \left(\frac{P}{69\ {\rm ms}}\right)^{7/6}\ {\rm G}$$

where f is a numerical factor ~0.5–1.0 depending on the detailed assumptions made and on whether the equations used are those for accretion from a wind[12,13], or for accretion through a disk[14]. We can envisage a situation where the magnetic field is close to this critical value and, as periastron is approached, essentially no accretion is possible until a sufficiently large amount of accreting matter is available to produce a high X-ray luminosity. This would explain the very large on/off ratios for the X-ray flux, the sharp turn-ons and turn-offs, and the small duty cycle of the on state. In this picture although the neutron star would spin-down between outbursts it might either spin-up or spin-down during the outburst. The contribution to the short-term $\dot{P}$ would in either case be small compared with that observed. We note that if the magnetic field were too small non-uniform accretion onto the newton star surface, necessary to explain the 69-ms modulation of the X-ray flux, would not occur.

We may consider whether a mechanism analogous to Roche Lobe overflow can be responsible for the accretion. To estimate the size of the companion's critical lobe at periastron, we adapted the results quoted by Bahcall[15] to the pseudo-potential given by Avni[16] describing the gravitational/centrifugal forces on a test particle co-rotating with the companion and in an eccentric binary system. Using the parameters found by Charles *et al.* for the quiescent state ($M_{\rm opt} = 12 M_\odot$, $R = 12.6\, R_\odot$), we find that for critical-lobe overfilling $0.6 \leq e \leq 0.7$, depending on

whether the primary is not rotating, rotating at the mean orbital angular velocity, or rotating at the orbital velocity at periastron. An eccentricity in this range is consistent with the observed $\dot{P}$ if the line of apsides makes an angle of 55°–65° with the line of sight (assuming the observation took place at periastron).

In conclusion, the observed P, $\dot{P}$ imply that A0538−66 is an eccentric binary system in which a neutron star with modest magnetic field ($B \leq 10^{11}$ G) approaches so close to its B star companion that massive accretion takes place. The region of hot gas giving rise to the optical outburst must be comparable in size with the separation of the two stars, and surrounding one or both of them. These deductions are based on data lasting only 4,000 s and observations of pulsations over longer periods would yield a great deal of information about this unique system. We urge that attempts be made to observe corresponding pulsations at optical and radio wavelengths.

We thank J. Halpern for assistance with the MPC analysis and F. R. Harnden for analysis of HRI data. J.E.G. is partially supported by an A. P. Sloan Fellowship and grant NAS8-30751.

Received 10 March; accepted 19 April 1982.

1. White, N. E. & Carpenter, G. F. *Mon. Not. R. astr. Soc.* **183**, 11–15P (1978).
2. Johnston, M. D. *et al. Astrophys. J. Lett.* **230**, L11–14 (1979).
3. Skinner, G. K. *Nature* **288**, 141–143 (1980).
4. Skinner, G. K. *et al. Astrophys. J.* **240**, 619–627 (1980).
5. Pakull, M. & Parmar, A. *Astr. Astrophys.* **102**, L1–4 (1981).
6. Charles, P. A. *et al. Space Sci. Rev.* **30**, 423–431 (1981).
7. Skinner, G. K. *Space Sci. Rev.* **30**, 441–446 (1981).
8. Brown, R. L. & Gould, R. J. *Phys. Rev.* **D1**, 2252–2256 (1970).
9. Schrier, E. J. *et al. Astrophys. J.* **204**, 539–547 (1976).
10. Sadeh, D. *et al. Astrophys. J.* (in the press).
11. Charles, P. A. *et al. Mon. Not. R. astr. Soc.* (submitted).
12. Davidson, K. & Ostriker, J. P. *Astrophys. J.* **179**, 585–598 (1973).
13. Ghosh, P. & Lamb, F. K. *Astrophys. J.* **234**, 296–316 (1979).
14. Lamb, F. K. *et al. Astrophys. J.* **184**, 271–283 (1973).
15. Bahcall, J. N. *A. Rev. Astr. Astrophys.* **16**, 241–264 (1978).
16. Avni, Y. *Astrophys. J.* **209**, 574–577 (1976).

Evidence for two pre-Flandrian palaeosols in Buchan, north-east Scotland

E. Rodger Connell*, Kevin J. Edwards† & Adrian M. Hall‡

* Department of Geography, University of Aberdeen, Aberdeen AB9 2UF, UK
† Department of Geography, University of Birmingham, Birmingham B15 2TT, UK
‡ Department of Geography, University of St Andrews, St Andrews KY17 9AL, UK

Evidence is presented for the existence of two distinct pre-Flandrian palaeosols at a single site, Kirkhill Quarry, in Buchan, north-east Scotland. This is the first such site reported from Scotland and it is possible that the two palaeosols represent interglacial episodes which are equivalent in age to last interglacial (compare with Ipswichian) and the penultimate interglacial (compare with Hoxnian)—a combination so far unknown in the British Isles. The stratigraphy also demonstrates that the part of the Buchan plateau under study may have been glaciated on at least three separate occasions.

Kirkhill Quarry (NGR NK 012 529) is located 13 km north-west of Peterhead, Aberdeenshire (Fig. 1*a*). The surface of the site lies at ~75 m OD and the bedrock consists of metasediments of the Dalradian Supergroup[1] intruded by a complex acid/basic dyke within which the quarry was opened.

The most complete development of the Pleistocene stratigraphy is seen in the quarry's east and north faces (Fig. 1*b*) and it is the evidence from these that is used to interpret the history of the site. Preservation of the sequence is a result of the protection afforded by the irregular bedrock surface consisting of very steep-sided channels or basins. The topographic location of these features and their association with glacigenic sediments suggest an origin by glacial or glacifluvial erosion facilitated, in part, by the weathered nature of the basic dyke rock.

The earliest phase of sedimentation at the site is recorded by gravels and coarse sands of fluvial (perhaps glacifluvial) origin. Typically this deposit has a light olive-brown colour (2.5 Y 5/4)[2] but the upper zones of both profiles consist of a 19-cm thick light-grey (10 YR 7/1) horizon overlying 7 cm of greyish-brown (10 YR 5/2) coarse sand containing ~1.4% organic matter. The latter is most clearly represented as a single organic band (~2 cm thick) in the east face. The north face stratigraphy is more complex with two main organic bands; the lower is ~1 cm thick and occurs within a blocky rubble matrix and the upper is ~6 cm thick and lies within bands of grey/yellow sands. There is deformation of the grey sand/organic sections of both profiles and a frost crack is visible in the east face. These horizons are overlain by coarse angular felsite debris which has probably been produced by frost shattering of the adjacent channel walls. This material is succeeded by till.

The till is present in both north and east faces where it is made conspicuous by its yellowish-brown matrix colour (10 YR 5/4) and its content of soft friable igneous and metamorphic clasts. In the north face section, the upper 70 cm of till have a strong brown colour (7.5 YR 5/6) with coarse distinct yellowish-red (5 YR 4/6–5/8) and greyish-brown/brown (10 YR 5/2–5/3) mottles. This would appear to be a weathering profile (discussed below) and it also displays signs of cryoturbation with the erection of clasts.

The till is buried beneath a head deposit which averages 60 cm in thickness and this is succeeded by further till deposits representing the last major event, or events, recognized. This situation is of some complexity because tills of both north-western (north face till derived from Moray Firth Mesozoic clays) and western (east face till derived from the igneous and metamorphic rocks of western Buchan) provenances are present. The availability of sections has not yet made it possible to determine which of these tills is the older. In contrast to the

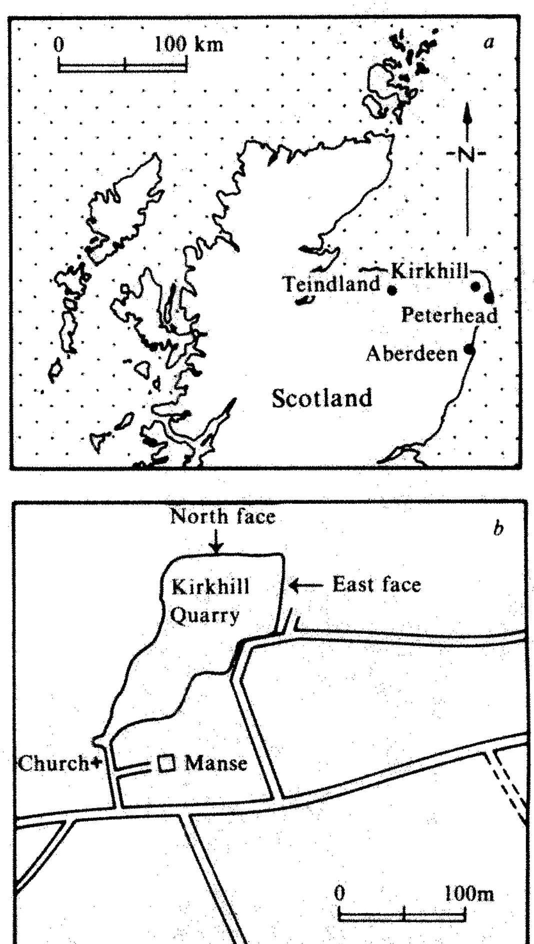


Fig. 1 *a*, Location of Kirkhill, Scotland. *b*, Kirkhill Quarry and location of sampling profiles.

Table 1 Summary of pollen data

Local pollen assemblages	No. of sampling levels North face	East face
Upper palaeosol		
KQ-6 *Alnus*	1	
Lower palaeosol		
KQ-5 *Calluna*		1
KQ-4 Gramineae—*Alnus*		2
KQ-3 Gramineae—*Alnus*-*Solidago*(+*Quercus*, *Picea*)	2	1
KQ-2 *Pinus*-*Alnus* (+*Tilia cordata*)		1
KQ-1 *Betula*—coryloid (+*Ulmus*, *Quercus*, *Picea*)		1

underlying and apparently severely weathered till, the upper tills have a fresh unweathered appearance and contain hard intact clasts.

The north face till is overlain by glacifluvial sand which fills a shallow channel. This sand has a humus iron podsol developed in it and appears to have been partially truncated by quarrying activity at some locations. Outside the channel limits, gleyed brown earths have developed in the upper tills.

Weathering in the light grey basal sands has resulted in the bleaching and softening of felsite clasts and the almost complete removal of such minerals as augite, hornblende and biotite from the fine sand fraction. Such a sequence of horizons resembles podsol-type soil profiles which have undergone weathering and translocation of iron and organic matter in a humid temperate climate.

The condition of the soft, friable clasts in the lower till would have precluded transportation and so a period of subaerial weathering after till deposition is indicated. In the north face weathered section of this till, clay mica has been altered to interstratified mica-smectite. X-ray diffraction analyses have identified the presence of lepidocrocite, an iron oxyhydroxide found in hydromorphic soils of temperate humid climate[3]. In thin section, mottles are seen as areas of iron oxide segregation and evidence of clay translocation is provided by numerous large angular papules of strongly orientated clay minerals, interpreted as void argillans[4] disrupted by subsequent cryoperturbation[5]. Taken together, these features indicate both gleying and translocation of clay, processes that are characteristic of B horizons of certain soil groups[6]. The upper 70 cm of the weathered till is thus interpreted as the truncated remnant of a soil formed under a humid temperate interglacial climate. The severity of clast weathering supports this interpretation.

Sediment samples from the basal and upper section of the north and east face profiles were prepared for pollen analysis by standard acetolysis and hydrofluoric acid procedures[7]. Sampling of organic layers for macroscopic plant and animal remains produced nothing. Only nine sampling levels produced sufficient pollen to permit useful discussion of the contained microfossils. Some of the spectra came from parallel sampling profiles and it was considered possible to place them in an internally consistent sequence by reference to lithostratigraphy and pollen content. The results are summarized in Table 1 and sample levels are indicated in Fig. 2. Pollen samples from the lower soil profile display some of the sharp inter-spectral changes found in the early buried podsol at nearby Teindland[8] and do not suggest the homogenization of assemblages which can result from microfaunal mixing or downflow of pollen through a permeable soil profile[9,10] (though such processes may still have occurred in part). The pollen assemblages have a mean arboreal representation (including shrubs) of 40.3% of total land pollen with a range of from 8.9% for *Calluna* assemblage KQ-5 to 96.1% for *Pinus-Alnus* assemblage KQ-2. The sequence KQ-1 to KQ-5 may well represent the relict registration of a vegetational landscape passing through an interglacial cycle[11,12] with a major local woodland component around the Kirkhill site. It is not possible to say whether the low amounts (1.0–5.0%) of such thermophilous taxa as *Ulmus* and *Quercus* in KQ-1 and KQ-3, *Tilia cordata* in KQ-2, as well as *Picea* in KQ-1 and KQ-3, indicate the local presence of elm, oak, lime and spruce or whether their pollen grains reached the site by long-distance wind transport from warmer areas to the south or east. If these taxa are of local origin (and not derived secondary pollen), their presence strongly supports an interglacial interpretation for the basal palaeosol.

A single sediment sample from the weathered section of the north face profile produced a pollen spectrum (KQ-6) dominated by *Alnus* (73.9% of the total land pollen) together with lower frequencies of *Betula* (4.3%) and coryloid (8.7%) grains. Such an assemblage would not be inconsistent with an interglacial phase.

Note that all pollen preparations contained microfragments of charcoal, while sieving for animal and plant remains also produced charcoal macrofragments in the basal organic levels. While evidence of burning is not limited to warm or temperate environments, it is more likely that such environments would provide a suitable context for the occurrence of natural fires.

Radiocarbon assay of three samples of organic material apparently representing both *in situ* soil (east face) and redeposited material (upper organic layer, north face) produced ^{14}C dates within the age range $33{,}810\pm^{630}_{590}$ (SRR-1636) to $45{,}630\pm^{1{,}740}_{1{,}430}$ (SRR-1635) yr b.p. (Fig. 2 and Table 2). The association of these ostensibly middle Devensian dates with pollen of a relatively thermophilous nature and linked with the stratigraphical context of this buried soil is considered anomalous. These dates must be regarded as minimal estimates of the true sample ages. This situation has been encountered with other Scottish interglacial sites and is generally thought to reflect sample contamination[13,14]. The $\delta^{13}C$ values lie within the normal range for terrestrial organic materials from European localities.

The sequence of events inferred from Kirkhill Quarry can be briefly summarized as: (1) erosion of channels/basins and deposition of fluvial or glacifluvial sands and gravels; (2) weathering and pedogenesis in humid temperate interglacial conditions to produce a podsol-like soil profile in the sands and gravels; (3) disturbance and burial of the podsolic profile by periglacial sediments followed by deposition of till; (4) weathering and soil development in the till under a second period of

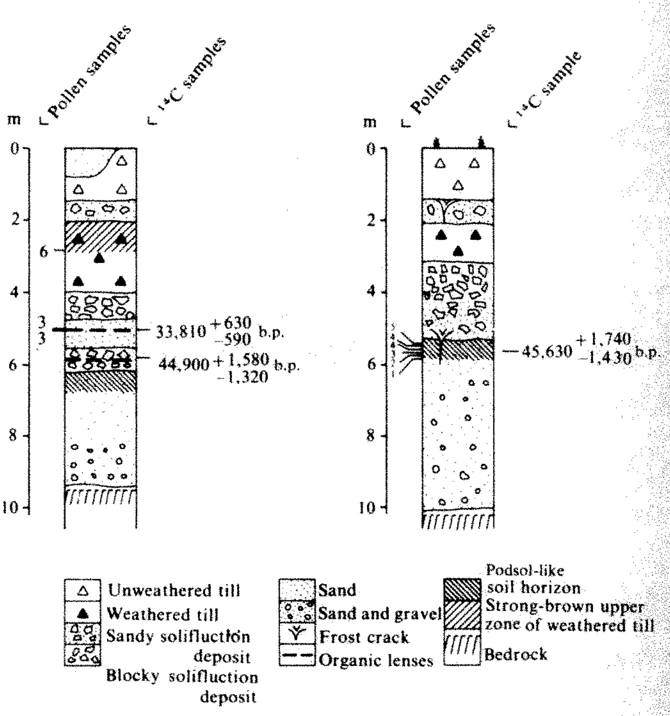


Fig. 2 Stratigraphic columns and positions of sampling levels for palynology (numbers indicate KQ-assemblage designations) and radiocarbon dating.

Table 2 Radiocarbon dates from Kirkhill Quarry

Sample no.	Location	Depth below surface (m)	Date (yr b.p.)	$\delta^{13}C$ value
SRR-1635	East face	5.69–571	$45,630 \pm ^{1,740}_{1,430}$	−26.3%
SRR-1636	North face	5.06–5.12	$33,810 \pm ^{630}_{590}$	−27.0%
SRR-1637	North face	5.895–5.905	$44,900 \pm ^{1,580}_{1,320}$	−25.7%

humid temperate climate; (5) periglacial disturbance of the soil profile and final glaciation of the site resulting in the deposition of the fresh, unweathered, upper tills; (6) erosion of channel and deposition of glacifluvial sands (north face); (7) Flandrian pedogenesis in tills and sands.

The Kirkhill Quarry sequence is important for several reasons. It is only the sixth polleniferous pre-late Devensian site known from Scotland and only the fourth claiming interglacial status[8,12,15–17]. The site is unique thus far in Scotland in providing evidence for two interglacial phases whose individual identity can be demonstrated in stratigraphical superposition within a single section. Kirkhill Quarry lies within the so-called 'Moraineless Buchan'[18,19] and outside the limits of the Moray Firth–Strathmore glaciation. As the latter is widely regarded as being of late Devensian age[20] it is possible that the upper tills at the site are even older. Despite this problem, the unweathered nature of these upper tills encourages the view that they date to a phase of Devensian glaciation. If this is the case then the simplest interpretation for the underlying interglacial horizons is that they are last interglacial (compare with Ipswichian) and penultimate interglacial (compare with Hoxnian) in age and the Kirkhill Quarry sequence would then be the first site for which this combination of deposits has been reported from the British Isles. The palaeofloras, whilst suggesting interglacial conditions, do not permit obvious correlation with the varied interglacial deposits from further afield[12,16,21–24]. The uppermost *Alnus*-dominated spectrum KQ-6 at Kirkhill may be compared with the *Alnus*–Gramineae–*Plantago lanceolata* pollen assemblages zone T-1 at nearby Teindland[8] which was assigned to the Ipswichian interglacial, and the *Alnus* zone ScB2 at Scandal Beck, Cumbria[25], which was tentatively referred to the Ipswichian partly by reference to the Teindland data. The Gramineae–*Alnus*–*Solidago* assemblage KQ-3 in the lower palaeosol at Kirkhill also resembles Teindland spectrum T-1, and assuming that two distinct pre-Flandrian interglacial phases are represented at Kirkhill then this may indicate the cyclical nature of interglacial vegetational change.

If the above reasoning is accepted, then the age of the upper palaeosol would be no younger than last interglacial. If it should prove to be older then the site becomes even more unusual in the Scottish context as the area possesses no known interglacial deposits earlier than the Hoxnian date claimed for Fugla Ness, Shetland[16]. If the upper palaeosol at Kirkhill is of Hoxnian age (and assuming no hiatuses), the basal podsol would chronologically be of Cromerian age and a profile combining both Hoxnian and Cromerian stage deposits is similarly unknown from the British Isles. If it is accepted that the two palaeosols represent distinct interglacial events and if the basal sands and gravels are glacifluvial in origin, then the stratigraphy demonstrates that this part of the Buchan plateau has been glaciated on at least three occasions. The earliest episode, indicated by the basal glacifluvial deposits, would be comparable in age with at least the Anglian glaciation of the English sequence[22].

We thank Banff and Buchan District Council for access to the site; Professor J. W. Parsons, Dr M. J. Wilson and Mr D. M. L. Duthie for chemical and mineralogical analyses; Dr D. D. Harkness for radiocarbon dates; Dr E. A. FitzPatrick for thin sections and advice; Dr P. C. Buckland who searched for animal remains; and Professor F. W. Shotton for comments on an earlier draft. E.R.C. acknowledges the receipt of a NERC research studentship.

Received 3 March; accepted 22 April 1982.

1. Harris, A. L. & Pitcher, W. S. in *A Correlation of Precambrian Rocks in the British Isles* (eds Harris, A. L. *et al.*) 52–75 (Geological Society of London, 1975).
2. *Munsell Soil Color Charts* (Munsell Color, Baltimore, 1975).
3. Schwertmann, U. & Taylor, R. M. in *Minerals in Soil Environments* (eds Dixon J. B. & Weed, S. B.) 145–180 (Soil Science Society of America, Madison, 1977).
4. Brewer, R. *Fabric and Mineral Analysis of Soils* (Krieger, New York, 1976).
5. Bullock, P. & Murphy, C. P. *Geoderma* **22,** 225–252 (1979).
6. Avery, B. W. *Soil Classification for England and Wales (Higher Categories)* (Soil Survey Technical Monogr. no. 14, Harpenden, 1980).
7. Faegri, K. & Iversen, J. *Textbook of Pollen Analysis* (Blackwell, Oxford, 1975).
8. Edwards, K. J. Caseldine, C. J. & Chester, D. K. *Nature* **264,** 742–744 (1976).
9. Dimbleby, G. W. *J. Soil Sci.* **12,** 1–11 (1961).
10. Havinga, A. J. *Geologie Mijnb.* **53,** 449–453 (1974).
11. Iversen, J. *Uppsala Univ. Årbok* **6,** 210–215 (1958).
12. Birks, H. J. B. & Peglar, S. M. *New Phytol.* **83,** 559–575 (1979).
13. Sissons, J. B. *Boreas* **10,** 1–17 (1981).
14. Shotton, F. W. *Geol. J.* **8,** 387–394 (1973).
15. FitzPatrick, E. A. *Nature* **207,** 621–622 (1965).
16. Birks, H. J. B. & Ransom, M. E. *New Phytol.* **68,** 777–796 (1969).
17. Edwards, K. J. & Connell, E. R. *Quat. Newslett.* **33,** 22–28 (1981).
18. Charlesworth, J. K. *Trans. R. Soc. Edinb.* **62,** 769–928 (1955).
19. Synge, F. M. *Scott. geogr. Mag.* **72,** 129–143 (1956).
20. Jardine, W. G. & Peacock, J. D. in *A Correlation of Quaternary Deposits in the British Isles* (eds Mitchell, G. F. *et al.*) 53–59 (Geological Society of London, 1973).
21. West, R. G. *New Phytol.* **85,** 571–622 (1980).
22. Mitchell, G. F., Penny, L. F., Shotton, F. W. & West, R. G. *A Correlation of Quaternary Deposits in the British Isles* (Geological Society of London, 1973).
23. Warren, W. P. *Geol. Surv. Ir. Bull.* **2,** 315–322 (1979).
24. Mangerud, J., Sønstegaard, E., Sejrup, H. P. & Haldorsen, S. *Boreas* **10,** 137–208 (1981).
25. Carter, P. A., Johnson, G. A. L. & Turner, J. *New Phytol.* **81,** 785–790 (1978).

Biological significance of surface flooding in warm-core ocean eddies

D. J. Tranter*, G. S. Leech* & D. J. Vaudrey†

* Division of Fisheries Research and
† Division of Oceanography, CSIRO Marine Laboratories,
PO Box 21, Cronulla, New South Wales 2230, Australia

Deep convective overturn has an important role[1] in enhancing the productivity of warm-core ocean eddies. An eddy (F) formed from the East Australian Current in early 1978 overwintered further to the south and developed a significant phytoplankton bloom the following spring. We describe here another process which tends to terminate such blooms by submerging winter water, relatively rich in nutrients, beneath poorer water of more recent origin.

Warm-core eddies shed by the Gulf Stream[2–4], Kuroshio[5–7] and East Australian Current[8–10] are islands of warm water interacting with a cooler atmosphere. In winter, the eddy surface cools and a deep mixed layer develops by convective overturn. This mixing process distributes nutrients uniformly throughout the isothermal layer (thermostad[11]), where they are available for photosynthesis; but if phytoplankton are mixed to depths as great as 200, 300 or even 400 m, their production is limited by the availability of light. With the onset of summer heating, a cap of warmer water forms, which isolates the winter water below and conserves the phytoplankton in the light. From that point onwards, the thermostad has a characteristic signature by means of which it can be tracked (G. R. Cresswell, work in preparation).

The properties of eddy J, which had overwintered north of Sydney in 1979, changed dramatically between September and October (Fig. 1). A warm surface cap (shaded) appeared in October with a nitrate concentration $<5 \mu g\, N\, l^{-1}$ (compared with $20 \mu g\, N\, l^{-1}$ the previous month), without appreciable increase in particulate organic matter (5–$15 \mu g\, N\, l^{-1}$). There was no spring bloom like that which had occurred in eddy F the previous year[1]. Where had all the nitrate gone?

In September, eddy J had a neighbour to the north-east, eddy K, whose winter thermostad was warmer (by ~1 °C) and shallower (by ~100 m) than that of J[12]. Figure 2 shows the temperature–salinity characteristics of the surface mixed layer of J in August, September and October compared with the

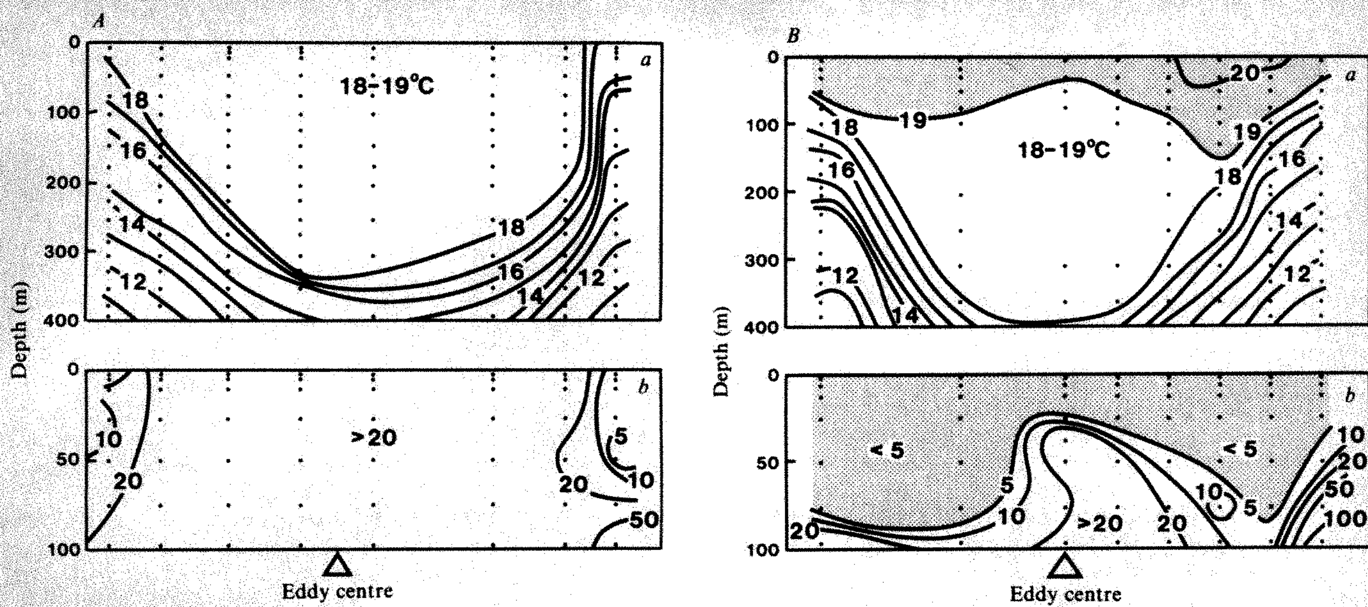


Fig. 1 Changes in the properties of eddy J between *A*, September and *B*, October 1979. *a*, Temperature (°C); *b*, inorganic nitrate (μg N l^{-1}). Eddy J developed a warm surface cap (shaded) and the nitrate concentration fell from >20 to <5 μg N l^{-1}.

September characteristics of K. For this purpose, the mixed layer was taken to be that stratum with a density range <0.1. The clusters of data points correspond to thermostads in the station profiles. Note that the J thermostad cooled by ~1 °C between August and September, an amount equivalent to the September differential between J and K. Note also that the October surface cap overlying the J thermostad had similar properties to the September thermostad of K. This warmer cap was evident in early October as an annulus around the outer edge of J and in late October as a surface blanket right across the eddy (G. R. Cresswell, work in preparation). We conclude that surface waters from eddy K spilled across the top of eddy J as an encircling ring which progressed towards its centre, eventually submerging it.

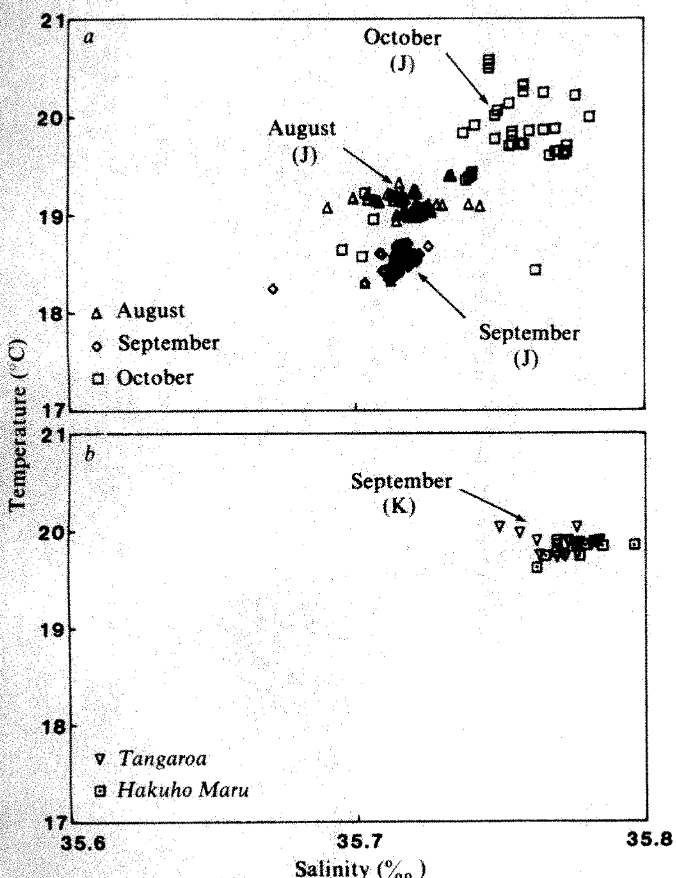


Fig. 2 Temperature–salinity characteristics of the surface mixed layer (where $\Delta\sigma t \not> 0.1$). *a*, J, August (SP9/79) September (SP10/79) and October (SP11/79). *b*, K, September (*Tangaroa* 1097, *Hakuho Maru* KH79-4). Note that the J thermostad cooled by ~1 °C before it was capped by a warm saline layer with similar properties to the K thermostad nearby.

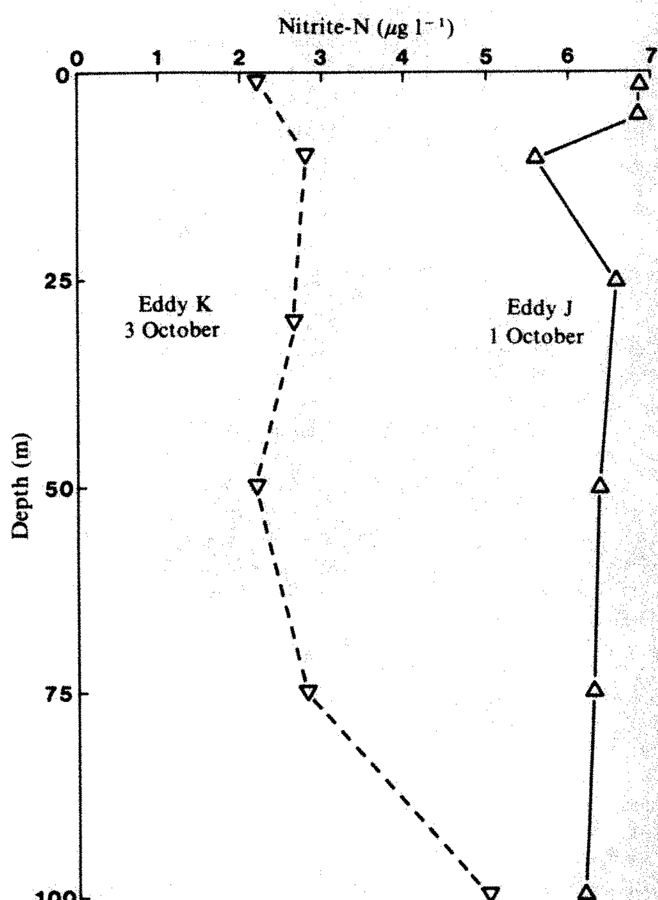


Fig. 3 Nitrite profiles through the K thermostad on 3 October (KH79-4 Stn 11 31°23′S 154°53.5′E); and the J thermostad on 1 October (SP10/79/238 33°20′S 152°46′E).

Tomosada[7] observed the crest of the Kuroshio to over-ride adjacent eddies and submerge the winter thermostad. Eddy B was over-ridden by the East Australian Current some time between October 1977 and February 1978[10]. The relevance of submergence in the present context is in relation to its impact on the productivity of the system. Nutrients elevated in winter by convective overturn of the water column are submerged again by overflow and consequently are not utilized in primary production. The subsequent productivity of eddy J was therefore determined largely by the nutrients of the overflow from K. There are no September nitrate data available for K, but its nitrite concentration was lower than that of J (Fig. 3), perhaps because of shallower winter mixing. If Nilsson and Cresswell are correct in their assertion that most of the warm-core eddies shed by the East Australian Current are reabsorbed at a later date, then we suggest that significant eddy productivity may occur only among those few that move permanently beyond its reach.

Dr Basil Stanton of the New Zealand Oceanographic Institute made temperature–salinity data available to us from *Tangaroa* cruise 1097 and Dr N. Taga of the Ocean Research Institute provided nitrite data from *Hakuho Maru* cruise KH 79-4. We thank Drs J. S. Godfrey and G. R. Cresswell for helpful discussions.

Received 27 January; accepted 21 April 1982.

1. Tranter, D. J., Parker, R. R. & Cresswell, G. R. *Nature* **284**, 540–542 (1980).
2. Saunders, P. M. *Deep-Sea Res.* **18**, 1207–1219 (1971).
3. Richardson, P. L., Cheney, R. E. & Worthington, L. V. *J. geophys. Res.* **83**, 6136–6144 (1978).
4. Hulliwell, G. R. & Mooers, C. N. K. *J. geophys. Res.* **84**, 7707–7725 (1979).
5. Hata, K. *Oceanogr. Mag.* **21**, 13–29 (1969).
6. Cheney, R. E. *J. geophys. Res.* **82**, 5459–5468 (1977).
7. Tomosada, A. *Bull. Tokai reg. Fish. Res. Lab.* **94**, 59–102 (1978).
8. Andrews, J. C. & Scully-Power, P. *J. phys. Oceanogr.* **6**, 756–765 (1976).
9. Nilsson, C. S., Andrews, J. C. & Scully-Power, P. *J. phys. Oceanogr.* **7**, 659–669 (1977).
10. Nilsson, C. S. & Cresswell, G. R. *Prog. Oceanogr.* **9**, 133–183 (1980).
11. Seitz, R. C. *J. mar. Res.* **25**, 203 (1967).
12. Stanton, B. R. *N.Z. J. mar. Freshwat. Res.* **15**, 289–297 (1981).

New basal Namurian (Upper Carboniferous) fishes and crustaceans found near Glasgow

S. P. Wood

Hunterian Museum, University of Glasgow, Glasgow G12 8QQ, UK

I report here a recent discovery of new species of sharks, together with the most complete and best-preserved Carboniferous shark known, in Bearsden near Glasgow. Sharks of this age are significant because of their high diversification rate during the Carboniferous. This, coupled with their poor fossil record, has led to "almost as many classifications of Palaeozoic chondrichthyan fish as there are taxa in this group represented by adequate specimens"[1]. The new shark discoveries also throw considerable morphological light on relatively fragmentary material collected during the past few years from the Upper Mississippian (Lower Namurian) of Montana[2]. The fish found also include 11 new palaeoniscid (early ray-finned) fishes which represent the first known British Namurian marine species. The site has also yielded a crustacean assemblage which is probably the best preserved from Europe or even the Northern Hemisphere. The only comparable crustacean fauna, from beds of Wesphalian age at Mazon Creek in Illinois, is larger but is mainly preserved as moulds whereas the Bearsden material preserves the cuticle intact. The first (and last) major discovery of a substantial vertebrate and crustacean marine fauna from the British Carboniferous was made in 1879 when the Glencartholm site (Dumfries and Galloway) was found by Macconochie of the Scottish Geological Survey[3] in the Lower Carboniferous, Upper Border Group (Upper Dinantian, Asbian).

The Bearsden fossils are contained in a series of beds of marine shale which vary in lithology and have minor non-marine intercalations. Some specimens were collected from weathered exposures along the Manse Burn but most were obtained from a temporary quarry 5×3 m in area excavated in the south bank. This quarry is marked locality 1 on a detailed map at the NCC Geological Review Unit, Newbury to which professional enquiries should be directed. The *Crangopsis* marker band (Fig. 1) outcrops twice in the burn and is also included in this map together with the general dip and structure of beds A–E.

From an examination of the shelly fauna, R. B. Wilson (personal communication) places the Bearsden fauna at the horizon of the Top Hosie Limestone. Currie[4] placed the base of the Pendleian, the lowermost Namurian stage, just below this limestone.

Specimens collected during the excavation at Bearsden, which took place throughout summer 1981, have been deposited at the Hunterian Museum of the University of Glasgow (HM V) and the Royal Scottish Museum, Edinburgh (RSM GY).

Outside North America, Carboniferous marine sharks have only been described from Glencartholm near Langholm[5,6] whilst non-marine sharks are known exclusively from the Edinburgh area[6–10]. Both are of Asbian age in the Dinantian of Scotland[11]. In addition, xenacanths have been recorded from the Permo-Carboniferous of Europe[12].

None of the marine sharks discovered at Bearsden occurs at the Edinburgh or Glencartholm sites, and the non-marine

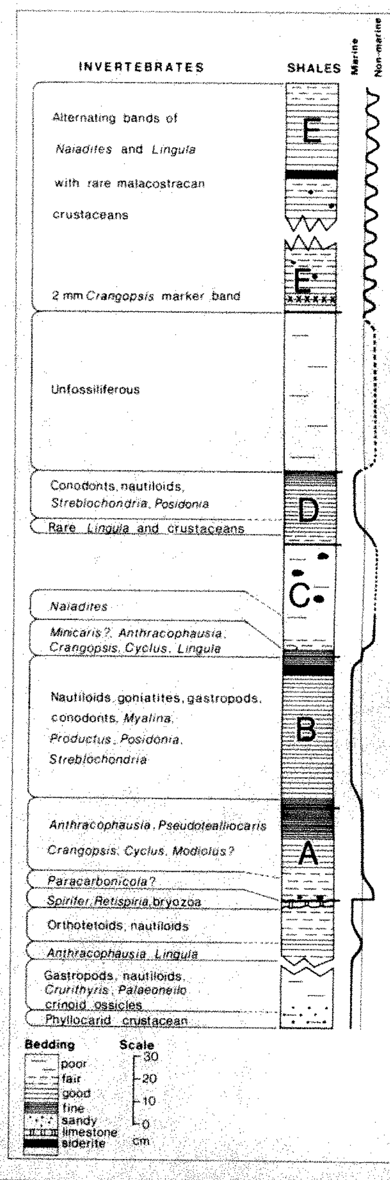


Fig. 1 Measured section in Manse Burn area. Lithological horizons used in the text are all approximate equivalents of Top Hosie limestone in Scotland.

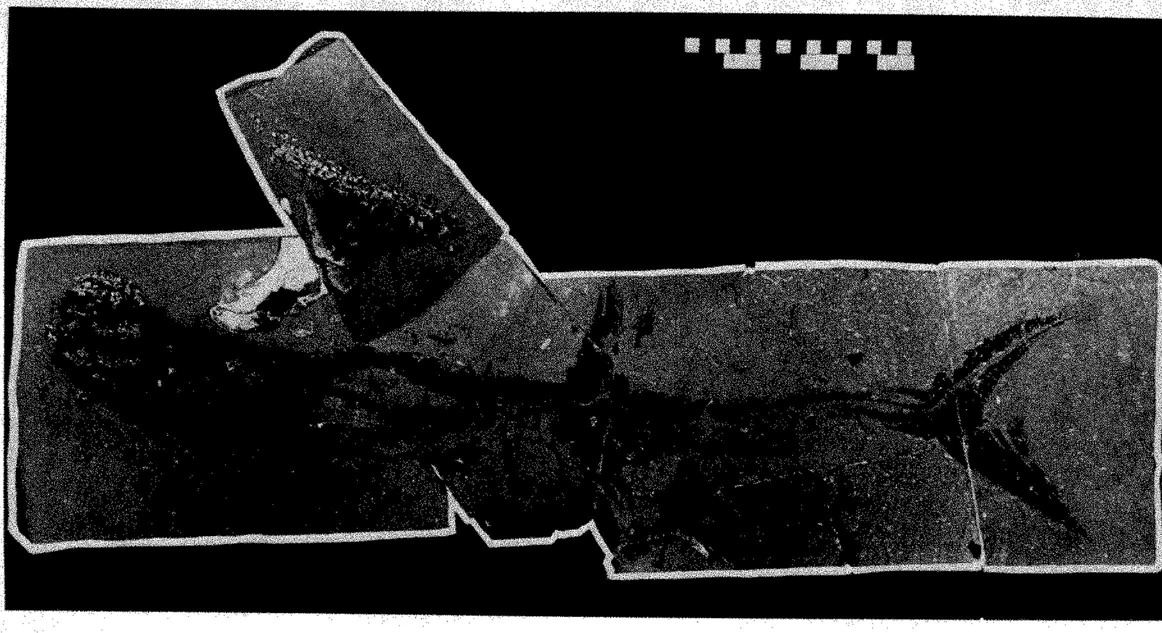


Fig. 2 Lateral view of superbly preserved anacanthous male shark (HM V8246). Photographed in methanol; upper scale bar in cm.

phalacanthous *Tristychius arcuatus* Agassiz is present at Bearsden and Edinburgh, but not at Glencartholm, despite many references to the contrary[7]. On the other hand, the only fish found at both, the phalacanthous *Onychoselache traquairi* Dick, is apparently absent at Bearsden. The Bearsden sharks are probably the best-preserved cartilaginous fish known from the Palaeozoic. One of them, HM V8246 (Fig. 2), should help to clarify the unsatisfactory systematics of anacanthous sharks. It could be *Cladodus neilsoni* Traquair, at present known only from the anterior end of a skeleton of similar age from East Kilbride and described by Traquair in the nineteenth century[13], or equally the Bear Gulch *Stethacanthus altonensis* (St John and Worthen)[1,14].

HM V8246 has a peculiar structure posteriorly attached to the dorsal spine near its base. It is well calcified, thick, apparently inflexible and triangular in shape. The posterio-dorsal margin is armoured with formidable monocuspid tooth-like denticles which increase in size posteriorly. Similar, though smaller, denticles form an anteriorly positioned 'helmet' on the head whilst behind the pelvic fins, external sexual organs identify the individual as male. If restored parallel to the body, the long axis posteriorly attached to each pectoral fin extends backwards well beyond the pelvic fins. These axes may have been used to grasp the female during copulation. The grotesque dorsal armament may also have had a sexual role as some extant forms, when excited, make extensive wounds in the flanks of females during foreplay.

S. altonensis and *C. neilsoni* may be synonymous as study of the type specimen of the latter (RSM 1911.62.52) suggests that the two posterior basal plate radials of the right pectoral fin are damaged, and it is the drifted distal portions lying against the axis cartilages which Traquair misinterpreted as axial radials. His interpretation has since been used by Moy-Thomas and Miles[15] and by Romer[16] when comparing reconstructed pectoral fins of other Palaeozoic sharks. The realization that the post-metapterygial axis of *C. neilsoni* is free of radials and that this condition may also be true for *S. altonensis* supports their being synonymous.

In RSM GY 1981.63.23, another Bearsden specimen of the same species, the dorsal spine and basal plate are exposed in cross-section in juxtaposition and the articulation of these elements is no different from that illustrated by Maisey[17] for the Upper Devonian shark *Cladoselache*. Maisey suggests that the unornamented dorsal spine of *Cladoselache* was deeply inserted. However, the undisturbed anatomy of HM V8246, a shark belonging to the same anacanthous group, contains an unornamented spine of similar design which is in 'life' position and is clearly shallow based. It may therefore be significant that shallow emplacement of the dorsal spine in *Cladoselache* was formerly suggested by Harris[18].

Additional Bearsden material including an isolated head (HM V8250) suggests an overall length of up to 1.3 m for *C. neilsoni* (=*S. altenensis?*). Lund (personal communication) considers HM V8250 to be female and he independently suspects synonomy between *Cladodus* and *Stethacanthus*, the latter being senior. Zangerl[19] noted the possibility of a generic relationship between the Westphalian *Symmorium* and *C. neilsoni*, and the newly discovered Bearsden material reinforces this suggestion, although the peculiar anterior dorsal structure is apparently absent in the former. Another important shark in the collection is only 12.5 cm long, has small cladodont teeth, lacks a fin spine and displays a series of distal radials above the vertebral column, suggesting the retention of a long-based dorsal fin as in *Chondrenchelys*. However, the chordacentra of the latter are absent. The poorly known bradyodont group is represented at Bearsden by a single complete suite of crushing teeth.

Eleven new palaeoniscid species seem to be represented. Many of the specimens are complete and all are well preserved. One aberrant form will require a new suprageneric taxon. This small fish has a large head, a raised denticle centrally positioned on each scale and long tapering projections extending beyond the pectoral and caudal fins, the former consisting of modified fulcral scales. The others collected include new species of *Elonichthys*, *Rhadinichthys* and *Mesopoma* (the latter identified by K. Lowney). In addition to the palaeoniscids, the deep-bodied chirodontid *Chirodus crassus* Traquair (Fig. 3) is present as several fine specimens which will allow accurate redescription, especially of the pelvic fins, hitherto presumed absent[20].

Acanthodians are represented by *Acanthodes sulcatus* Ag. The material includes a superb example (HM V8251), whilst in another specimen the first-known Carboniferous acanthodian braincase was recognized by A. L. Panchen and T. R. Smithson.

The sarcopterygians are poorly represented: dipnoans are so far absent, the only coelacanth seen is an isolated head, whilst the other crossopterygians occur as rare rhizodont and osteolepid fragments.

Another remarkable feature of the Bearsden discovery is the superbly preserved crustacean fauna, rivalling that from any British or European locality. It complements the well-known Westphalian fauna of Mazon Creek, Illinois, but is less diverse

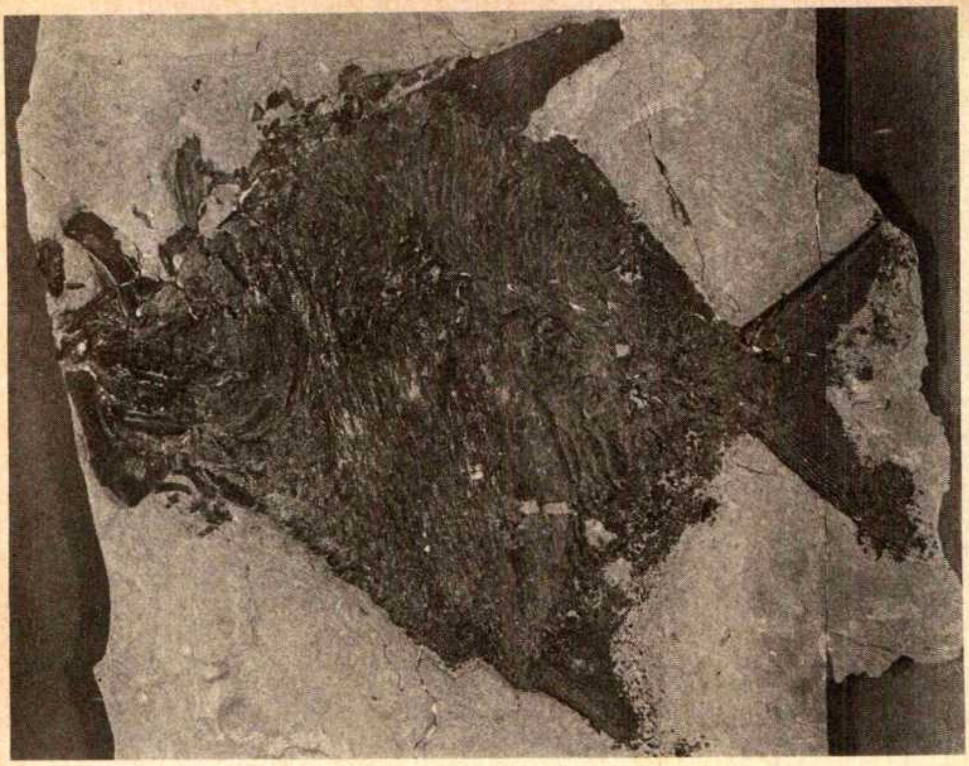


Fig. 3 *Chirodus crassus* (HM V8247), a deep-bodied palaeoniscoid 31 cm long, yielding new anatomical detail. Photographed in methanol.

than the North American fauna and consists of the malacostracans *Anthracophausia*, *Crangopsis*, *Pseudotealliocaris* and *Minicaris*?, together with the problematical '*Cyclus*'. This assemblage contains the near-shore and brackish water elements of the four marine communities proposed by Schram[21] which occur together, indicating a broader ecological niche than has been appreciated. Fine ornament and delicate structures are preserved in the cuticle of *Anthracophausia*, and W. D. I. Rolfe identified the raptorial appendages (Fig. 4) as the second antennae. This conclusion challenges the filter mode of feeding previously ascribed to *Anthracophausia*[22]. The occurrence of *Minicaris*? is noteworthy as this fossil is known from a single example in the Institute of Geological Sciences, Edinburgh, from a core sample of Asbian (middle Visean) age near Livingstone, Lothian Region[23].

Figure 1 shows the vertical distribution of the molluscs and some other elements of the fauna, and a relative salinity curve is drawn, using the invertebrates as indicators. Biofacies and lithofacies do not conveniently correspond, the former occasionally lagging behind the latter. The salinity profile

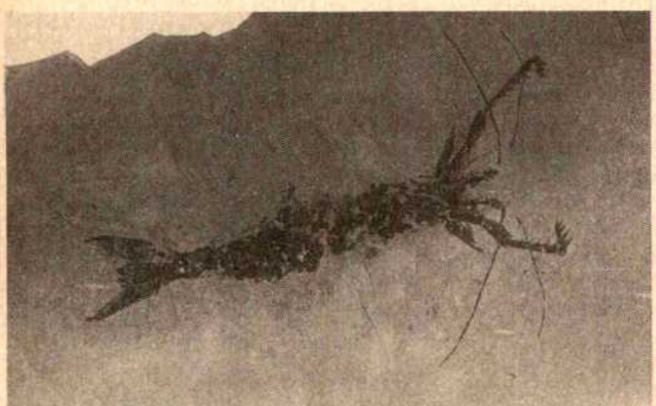


Fig. 4 *Anthracophausia dunsiana* (HM A2359) in dorsal view, an 8 cm long malacostracan crustacean. The anterior appendages were hitherto unknown in this animal.

through these overlap periods is therefore tentative. There are also constraints (except on *Acanthodes*) in fish distribution, some palaeoniscoids occurring only in the marine phases of B and D beds, whereas others are restricted to the transitional stratigraphy of basal B and upper A and so on. Similarly, shark restriction is apparent: *Cladodus* is only found in beds B and D and *Tristychius* in beds A (middle) and C, the latter probably representing Dick's[10] non-marine *Diplodoselache* fauna typical of the Oil Shales around Edinburgh. Further sampling should clarify this picture.

The marine palaeoniscoid fauna at Bearsden is understandably different from the well-known non-marine species of the Scottish Midland Valley but as they also bear no specific relationship to the marine palaeoniscoids from Glencartholm, or the South African Witteberg assemblage[24], this raises interesting palaeoenvironmental questions.

In addition to the fish and crustaceans the collection contains microfauna and problematica and flora among which *Lepidophloios*, *Spathulopteris*, *Sphenopteris* and *Rhacopteris* have been recognized.

I thank the Leisure and Recreation Committee of Bearsden and Milngavie District Council for permission to quarry, Hunterian Museum staff for field assistance, Drs W. D. I. Rolfe, J. K. Ingham and A. L. Panchen for discussion, Dr R. B. Wilson for identifying and evaluating the molluscs, and Mr C. Cleal for identifying the flora. I acknowledge the help of Mr D. Barbour (drafting), Mr P. McGrail (typing) and Mr G. Grant (photography). I especially thank the Manpower Services Commission and Nature Conservancy Council for funding.

Received 16 February; accepted 31 March 1982.

1. Lund, R. *Ann. Carneg. Mus.* **45** (8), 161 (1974).
2. Melton, W. G. Jr *Proc. N. Am. Paleont. Conv.* **1**, 1202 (1971).
3. Traquair, R. H. *Trans. R. Soc. Edinb.* **30**, 15 (1881).
4. Currie, E. D. *Trans. R. Soc. Edinb.* **62**, 527 (1954).
5. Moy-Thomas, J. A. *Proc. zool. Soc. Lond.* **106**, 761 (1936).
6. Dick, J. R. F. & Maisey, J. G. *Palaeontology* **23**, 363 (1980).
7. Traquair, R. H. *Proc. Trans. R. Soc. Edinb.* **40**, 687 (1903).
8. Wood, S. P. *Scott. J. Geol.* **11**, 251 (1975).
9. Dick, J. R. F. *Trans. R. Soc. Edinb.* **70**, 63 (1978).
10. Dick, J. R. F. *Trans. R. Soc. Edinb.* **72**, 99 (1981).
11. George, T. N. *et al. Spec. Rep. geol. Soc. Lond.* **7**, 1 (1976).
12. Brongniart, C. *Bull. Soc. géol. Fr.* **3**, 16 (1888).
13. Traquair, R. H. *Trans. geol. Soc. Glasg.* **11**, 41 (1897).

14. Williams, M. E. thesis, Univ. Kansas (1979).
15. Moy-Thomas, J. A. & Miles, R. S. *Palaeozoic Fishes* 2nd edn (Chapman & Hall, London, 1971).
16. Romer, A. S. *Vertebrate Paleontology* 3rd edn (University of Chicago Press, 1966).
17. Maisey, J. G. *Neues Jb. Geol. Paläont. Mh.* **1977**, 47 (1977).
18. Harris, J. E. *Scient. Publs. Cleveland Mus. nat. Hist.* **8**, 1 (1938).
19. Zangerl, R. in *Interrelationships of Fishes* (eds Greenwood, P. H., Miles, R. S. & Patterson, C.) 1 (Academic, London, 1973).
20. Traquair, R. H. *Trans. R. Soc. Edinb.* **29**, 343 (1879).
21. Schram, F. R. *J. Paleont.* **55**, 1 (1981).
22. Schram, F. R. in *Mazon Creek Fossils* (ed. Nitecki, M. H.) 159 (Academic, New York, 1979).
23. Schram, F. R. *Fieldiana Geol.* **40**, 1 (1979).
24. Gardiner, B. G. *Zool. J. Linn. Soc.* **48**, 423 (1969).

Carbon isotope ratios of apatite from fossil bone cannot be used to reconstruct diets of animals

Margaret J. Schoeninger* & Michael J. DeNiro†

Department of Earth and Space Sciences, University of California, Los Angeles, California 90024, USA

The reconstruction of animals' diets from measurements of stable isotope levels in fossils relies on the fact that the $^{13}C/^{12}C$ ratio of animal carbon reflects the $^{13}C/^{12}C$ ratio of dietary carbon[1,2]. Two phases in fresh bone, collagen and the carbonate occurring in apatite, the predominant bone mineral, have isotopic ratios that are related to the $^{13}C/^{12}C$ ratio of the diet[1]. The isotopic ratios of both phases have been used to study the diets of extant animals[3,4]. Reconstruction of the diets of fossil animals using the isotopic method has been limited to analysis of collagen preserved in bone[5–8]. It has not been possible to use the $^{13}C/^{12}C$ ratios of carbon in the inorganic phase of fossil bone for dietary reconstruction because most fossil bones contain significant amounts of calcium carbonate, deposited after the animal's death, that contribute to the CO_2 evolved from the bone during acid hydrolysis[3,4]. However, Sullivan and Krueger[9] recently presented data which led them to conclude that the $^{13}C/^{12}C$ ratio of CO_2 extracted by acid hydrolysis from the apatite phase of fossil bone records information about an animal's diet. We have now determined the $^{13}C/^{12}C$ ratios of both the apatite phase and the collagen of 24 fossil animal and human bones, and our results indicate that the $^{13}C/^{12}C$ ratio of fossil bone apatite cannot be used for dietary reconstruction.

Sullivan and Krueger[9] treated fossil bone with acetic acid to remove any secondary calcium carbonate that might have been present. They found that the $^{13}C/^{12}C$ ratios of the apatite carbonate fractions, which resisted acetic acid dissolution, were linearly correlated with $^{13}C/^{12}C$ ratios of the collagen for fossil as well as fresh bones. Because the collagen $^{13}C/^{12}C$ ratio is related to the isotopic composition of an animal's diet[1], they concluded that the $^{13}C/^{12}C$ ratios of apatite from fossil bone can be used to reconstruct aspects of the diets of animals that lived in the past.

There is a serious complication with the proposal[9] that the $^{13}C/^{12}C$ ratios of apatite in fossil bone can be used to reconstruct diet. Comparisons of the ^{14}C content of apatite with the ^{14}C content of collagen from the same bone or of other carbon-containing materials (such as charcoal) excavated in conjunction with the bone in question indicate that the apatite carbon can undergo exchange with carbon encountered in the post-mortem environment, either in groundwater or in the atmosphere (see, for example, refs 10–13). These diagenetic isotopic exchange processes can also affect the $^{13}C/^{12}C$ ratios of apatite[4]. To illustrate the magnitude of the stable isotopic shifts that can occur during diagenesis, we present the results of analysis of several suites of fossil bones.

We determined the $^{13}C/^{12}C$ ratios of the apatite and collagen of 6 modern and 24 fossil bones. These data, along with the ages and collection sites of the bones, are given in Table 1. Collagen was extracted from powdered bone as described previously[5] and its $^{13}C/^{12}C$ ratios determined following combustion by a modification of the Stump and Frazer method[14,15]. The $^{13}C/^{12}C$ ratios of the CO_2 liberated from apatite by reaction with anhydrous phosphoric acid were determined on separate aliquots of powdered bone that had been soaked in 50:50 (v:v) glacial acetic acid:water for two days before oxidation of the organic matter using sodium hypochlorite[1]. X-ray diffraction analysis demonstrated that the secondary calcium carbonate present in the fossil bones was removed by the acetic acid treatment.

We present a plot of the apatite $\delta^{13}C$ values against the collagen $\delta^{13}C$ values in Fig. 1. The line in Fig. 1 represents the relationship between collagen and apatite $\delta^{13}C$ values given by Sullivan and Krueger (see Fig. 1 legend). The data points for the six modern bones all lie close to this line. On the other hand, the differences between the observed apatite $\delta^{13}C$ value and that expected from the collagen $\delta^{13}C$ value for the fossil bones, based on the Sullivan-Krueger relationship, range from +4.6‰ to −12.1‰. This observation indicates that exchange processes have shifted the apatite $\delta^{13}C$ values from their original values, towards either more positive or more negative values, by as much as 12‰. Such a shift is significant when one considers that the average difference between the $\delta^{13}C$ values of C_3 and C_4 plants is about 14‰[16], while that between aquatic and most terrestrial food sources is about 7‰[1,8]. Determination of the

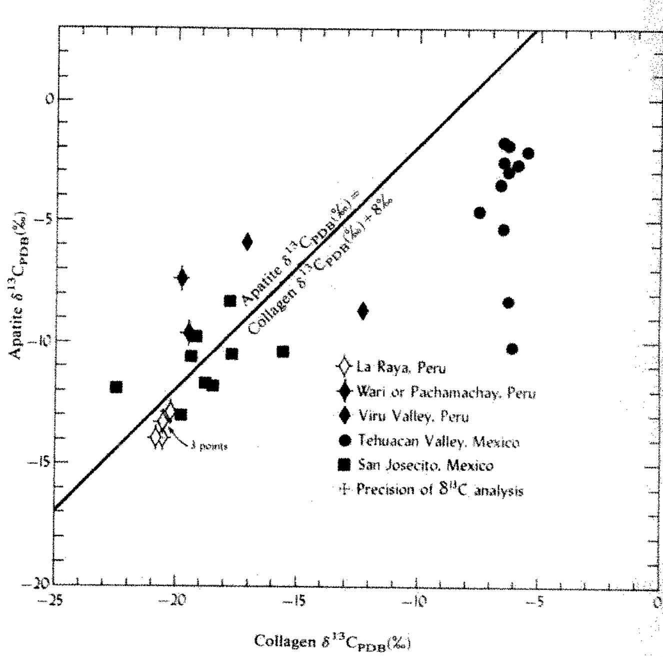


Fig. 1 The relationship between the $\delta^{13}C$ values of the apatite and collagen fractions of modern (open symbols) and fossil (closed symbols) bones analysed for this study. The line indicates the relationship given by Sullivan and Krueger[9]. Linear regression analysis of the data for fossil bones presented here produced the line: apatite $\delta^{13}C = 0.47$ collagen $\delta^{13}C - 1.29‰$ ($r = 0.80$). Sullivan and Krueger[9] indicated that apatite $\delta^{13}C$ = collagen $\delta^{13}C$ + 8‰. Linear regression analysis of their data yielded the line: apatite $\delta^{13}C = 1.07$ collagen $\delta^{13}C + 8.9‰$ ($r = 0.99$). The relationship given by Sullivan and Krueger is used in the discussion in the text. The values for $\delta^{13}C$ were calculated using the PDB belemnite carbonate as standard from the relationship:

$$\delta^{13}C = \left[\frac{(^{13}C/^{12}C)_{\mathrm{Sample}}}{(^{13}C/^{12}C)_{\mathrm{Standard}}} - 1\right] \times 1{,}000‰$$

* Permanent address: Department of Cell Biology and Anatomy, Johns Hopkins University School of Medicine, Baltimore, Maryland 21209, USA.
† Also Program in Archaeology.

0028-0836/82/240577—02$01.00

Table 1 Collection location, taxonomic identity, age and $\delta^{13}C_{PDB}$ values for the apatite and collagen fractions of the modern and fossil bones analysed for this study

Location	Species	Age (yr BP)	Apatite $\delta^{13}C_{PDB}$ (‰)	Collagen $\delta^{13}C_{PDB}$ (‰)
La Raya, Peru	*Lama pacos*	0	−13.3	−20.5
	Lama pacos	0	−12.9	−20.2
	Lama pacos	0	−14.0	−20.6
	Lama pacos	0	−14.0	−20.8
	Lama pacos	0	−13.3	−20.5
	Lama pacos	0	−13.3	−20.5
Pachamachay, Peru	*Lama glama*	4,500–3,750	−9.7	−19.5
San Josecito, Mexico	*Canis diris*	~20,000	−10.4	−15.6
	Cervus sp.	~20,000	−10.6	−19.4
	Cervus sp.	~20,000	−9.9	−19.2
	Equus sp.	~20,000	−11.9	−22.7
	Equus sp.	~20,000	−13.0	−19.8
	Felis concolor	~20,000	−11.8	−18.5
	Felis concolor	~20,000	−11.7	−18.8
	Ursus arctos	~20,000	−10.5	−17.7
	Ursus arctos	~20,000	−8.3	−17.8
Tehuacan Valley, Mexico	*Homo sapiens*	1,300–500	−8.3	−6.3
	Homo sapiens	1,300–500	−2.2	−5.5
	Homo sapiens	1,300–500	−2.7	−5.9
	Homo sapiens	1,300–500	−1.9	−6.3
	Homo sapiens	1,300–500	−3.5	−6.6
	Homo sapiens	1,300–500	−3.0	−6.3
	Homo sapiens	1,300–500	−10.2	−6.1
	Homo sapiens	1,300–500	−5.3	−6.5
	Homo sapiens	2,800–2,150	−1.8	−6.5
	Homo sapiens	2,800–2,150	−4.6	−7.5
	Homo sapiens	7,000–5,400	−2.6	−6.5
Viru Valley, Peru	*Homo sapiens*	2,200–1,500	−8.7	−12.3
	Homo sapiens	2,200–1,500	−5.9	−17.1
Wari, Peru	*Lama glama*	1,350–1,160	−7.4	−19.8

The ages given for the bones are generally based on dates obtained for other materials excavated with them and hence provide minimum and maximum limits on the actual ages.

relative amounts of C_3 versus C_4 plants or of aquatic versus terrestrial foods in the diet are the two applications of the isotopic method of dietary analysis that have been made to date[5–8]. Uncertainties in the estimates of diet $\delta^{13}C$ values on the order of 5–10‰, resulting from diagenetic exchange processes, would thus make dietary reconstruction based on fossil bone apatite $\delta^{13}C$ values completely unreliable.

Our treatment of the data reported here, as well as that used by Sullivan and Krueger[9], is based on the assumption that the fossil bone collagen $\delta^{13}C$ values have not been altered by diagenetic processes. This may not be a valid assumption[3–5]. However, some of the apatite $\delta^{13}C$ values presented in Table 1 can be interpreted without resorting to comparisons with the corresponding collagen $\delta^{13}C$ values. Eight of the 11 samples from the Tehuacan Valley come from a single period of the occupation of the site[5]. There is no evidence to suggest differences in status among these individuals[17,18]. Accordingly, it is reasonable to expect that their diets, and consequently the bone isotope ratios, would be similar. The $\delta^{13}C$ values of collagen from these samples agree to within 1‰, which is consistent with this expectation. The $\delta^{15}N$ values of collagen, which have also been shown to be determined by diet, for these same eight samples have a range of only 1‰[5]. Note, however, that the apatite $\delta^{13}C$ values for these samples range from −1.9‰ to −10.2‰. These observations indicate that the apatite carbon in these bones has undergone post-mortem exchange. The large range of $\delta^{13}C$ values suggests that individual bones have exchanged with carbon sources of different $^{13}C/^{12}C$ ratios and/or have undergone different amounts of exchange.

The results presented here demonstrate that bone apatite carbonate undergoes exchange with carbon encountered after the death of the animal and that these diagenetic exchange processes can shift the apatite $\delta^{13}C$ values by at least as much as 12‰. Uncertainties of this magnitude in estimates of diet $\delta^{13}C$ values cannot be tolerated, since differences in the $\delta^{13}C$ values of different foodstuffs whose relative consumption is of interest are only 7–14‰. The apatite carbon in some fossil bones may not have undergone post-mortem exchange. However, until a method is developed to identify such isotopically unaltered apatite, we conclude that $\delta^{13}C$ values of fossil bone apatite cannot be used to reconstruct ancient diets.

The following individuals supplied or helped us obtain fossil bones: M. West, J. E. Ericson, C. C. Patterson, E. Wing, R. S. MacNeish, J. Wheeler, A. Romano and D. P. Whistler. L. Armstrong and D. Winter performed the X-ray diffraction analyses and/or isotopic measurements. This work was supported by NSF grant BNS 79-24756.

Received 27 November 1981; accepted 2 April 1982.

1. DeNiro, M. J. & Epstein, S. *Geochim. cosmochim. Acta* **42**, 495–506 (1978).
2. Teeri, J. A. & Schoeller, D. A. *Oecologia* **39**, 197–200 (1979).
3. DeNiro, M. J. & Epstein, S. *Science* **201**, 906–908 (1978).
4. Land, L. S., Lundelius, E. L. Jr & Valastro, J., Jr *Palaeogeogr., Paleoclimatol., Paleoecol.* **32**, 143–151 (1980).
5. DeNiro, M. J. & Epstein, S. *Geochim. cosmochim. Acta* **45**, 341–351 (1981).
6. van der Merwe, N. J. & Vogel, J. C. *Nature* **276**, 815–816 (1978).
7. Burleigh, R. & Brothwell, D. *J. Archaeol. Sci.* **5**, 355–362 (1978).
8. Tauber, H. *Nature* **292**, 332–333 (1981).
9. Sullivan, C. H. & Krueger, H. W. *Nature* **292**, 333–335 (1981).
10. Hassan, A. A., Termine, J. D. & Haynes, C. V. Jr *Radiocarbon* **19**, 364–374 (1977).
11. Haas, H. & Banewicz, J. *Radiocarbon* **22**, 537–544 (1980).
12. Haynes, V. *Science* **161**, 687–688 (1968).
13. Tamers, M. A. & Pearson, F. J. Jr *Nature* **208**, 1053–1055 (1965).
14. Stump, R. K. & Frazer, J. W. *Nuclear Sci. Abs.* **28**, 746 (1973).
15. Northfelt, D. W., DeNiro, M. J. & Epstein, S. *Geochim. cosmochim. Acta* **45**, 1895–1898 (1981).
16. Smith, B. N. & Epstein, S. *Pl. Physiol.* **47**, 380–384 (1971).
17. Anderson, J. E. in *The Prehistory of the Tehuacan Valley* (ed. Byers, D. S.) Vol. 1, 91–113 (University of Texas Press, Austin, 1967).
18. MacNeish, R. S. in *The Prehistory of the Tehuacan Valley* (ed. Byers, D. S.) Vol. 1, 290–309 (University of Texas Press, Austin, 1967).

Oxygen production by endosymbiotic algae controls superoxide dismutase activity in their animal host

James A. Dykens & J. Malcolm Shick

Department of Zoology, University of Maine, Orono, Maine 04469, USA

Aerobic and aerotolerant organisms have evolved defenses against the toxic effects of molecular oxygen[1]. One protective mechanism involves the breakdown of the harmful superoxide radical by the enzyme superoxide dismutase (SOD). However, levels of oxygen elevated only slightly above normal atmospheric (P_{O_2} of 159 mm Hg) may overpower a cell's defense systems[2-4]. Although most animals do not naturally encounter oxygen pressures above 1 atm, hyperbaric oxygen levels normally occur in the tissues of marine animals that harbour intracellular algal symbionts, which in light generate more oxygen than is consumed by the combined host and symbionts[5-10]. We have now found that sea anemones *Anthopleura elegantissima* (Brandt) containing zooxanthellae (the symbiotic dinoflagellate *Symbiodinium microadriaticum*) in their gastrodermal tissues have SOD activities nearly two orders of magnitude greater than individuals totally lacking zooxanthellae.

Using oxygen microelectrodes we have recorded P_{O_2} values of up to 328 mm Hg in *A. elegantissima* gastrodermal tissues when the anemone is exposed to light intensities comparable with 33% of sunlight (Fig. 1). The bubbles that form on the surface of the anemones in bright light are likely to be composed of pure oxygen, as is the case in reef corals[7]. Thus, the host is subjected to a continuous flux of photosynthetically produced hyperbaric oxygen. Much attention has been paid to the mutualistic aspects of alga–invertebrate symbioses (such as transfer of photosynthetically fixed carbon compounds from algae to host in exchange for organic and inorganic metabolites)[11,12], yet little consideration has so far been given to the possibility of oxygen toxicity.

Molecular oxygen preferentially undergoes univalent reductions, one product of which is the highly reactive superoxide anion(O_2^-)[13]. Whether it is the superoxide anion *per se* or the hydroxyl radical (which may be produced by a reaction between O_2^- and H_2O_2) that is so destructive to cells remains debatable[14,15]. It is known, however, that O_2^- induces lipid peroxidation, depolymerizes hyaluronate, inactivates ribonuclease and destroys phagocytized bacteria when produced by activated polymorphonuclear leucocytes[16,17]. Virtually all aerobic organisms examined contain at least one form of the enzyme SOD[18] which catalyses the reaction $O_2^- + O_2^- + 2H^+ \rightarrow H_2O_2 + O_2$, and so prevents the cellular accumulation of O_2^- and its resultant damage.

We considered that *A. elegantissima* symbiotically associated with greater numbers of zooxanthellae would have proportionally higher levels of SOD activity, to cope with the higher rate of oxygen production. This has been confirmed: SOD activity was found to be positively correlated with chlorophyll content in symbiotic anemones (Fig. 2).

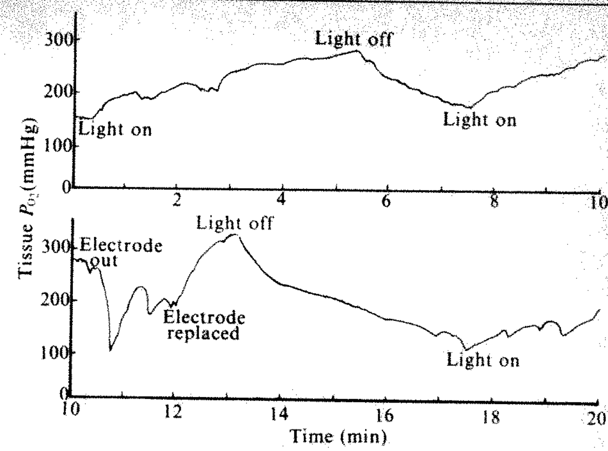


Fig. 1 Continuous record of changes in P_{O_2} in gastrodermal tissue of the zooxanthella-containing sea anemone *A. elegantissima* during illumination. Using a micromanipulator, a glass oxygen microelectrode (Transidyne General) was inserted into intact anemones at the base of the tentacles and illumination (315 µEinstein m^{-2} s^{-1}) was provided by high-intensity fibre optics. A few crystals of menthol were added to the seawater to reduce the anemone's movements so that the electrode would remain in position.

Anemones collected from the upper areas of the intertidal zone contain significantly ($P<0.01$) more chlorophyll than anemones collected from the lower intertidal ($\bar{x}$ = 5.21 µg per mg protein, s.d. = 2.28, n = 12 compared with 2.67 µg per mg protein, s.d. = 1.07, n = 17) (Fig. 2). Photosynthetically produced carbohydrates are an important source of nutrition in *A. elegantissima*[19,20]. As compensation for reduced feeding time, high intertidal anemones may rely more heavily on algal photosynthate than low intertidal individuals and would, therefore, maintain a greater photosynthetic capacity. Regardless, SOD activity per µg chlorophyll is identical in high intertidal ($\bar{x}$ = 2.42 U per µg chlorophyll, s.d. = 1.89, n = 12) and low intertidal anemones ($\bar{x}$ = 2.42 U per µg chlorophyll, s.d. = 1.74, n = 17), supporting the idea that SOD activity of the host is modulated by the amount of photosynthetically generated O_2 and O_2^- in the tissues. It has also been noted that *Euglena gracilis* raised phototrophically has almost three times the SOD activity of *E. gracilis* raised heterotrophically[21].

To examine further the relationship between photosynthetic activity and sea anemone SOD activity, we divided members of a single clone of high intertidal *A. elegantissima* into four groups, which were given different treatments as follows: (1) sunlight (control); (2) sunlight plus 10^{-5} M 3-(3,4-dichlorophenyl)-1,1-dimethylurea (DCMU) which inhibits photosynthesis; (3) darkness; and (4) darkness, plus constant exposure to 3 atm O_2. As expected, exposure to exogenous hyperbaric O_2 caused an increase in SOD activity of 62% (Table 1). Chlorophyll content tended to decrease (although not significantly) in both dark groups and in the DCMU-treated anemones (Table 1), in agreement with previous reports on the loss of zooxanthellae in both anemones and corals deprived of light[20,22]. The loss of photosynthetically produced O_2 by either

Table 1 Superoxide dismutase activity and chlorophyll content in *A. elegantissima*

	Sunlight (control)	Sunlight + 10^{-5} M DCMU	Darkness	Darkness + 3 atm O_2
SOD/protein (U mg^{-1})	3.81 ± 1.84*	1.41 ± 0.47†	2.90 ± 0.81*†	6.18 ± 1.71‡
Chl/protein (µg mg^{-1})	0.8423 ± 0.4630*	0.5424 ± 0.1452*	0.5406 ± 0.1961*	0.5233 ± 0.1662*

SOD activity and chlorophyll (Chl) content measured after 14 days in each of four treatments: 15 °C, n = 10 in each group, mean ± s.d. Means not significantly different by Tukey's test have same symbols. Control and DCMU-treated anemones were held outdoors in glass aquaria and exposed to at least 6 h direct sunlight (~860 µE m^{-2} s^{-1}) plus an additional 4–5 h of indirect sunlight (~680 µE m^{-2} s^{-1}) each day. Temperature (15 °C) was maintained with a thermostatted immersion refrigerator. Both dark-treated groups were held in a constant environment room at 15 °C.

0028-0836/82/240579—02$01.00

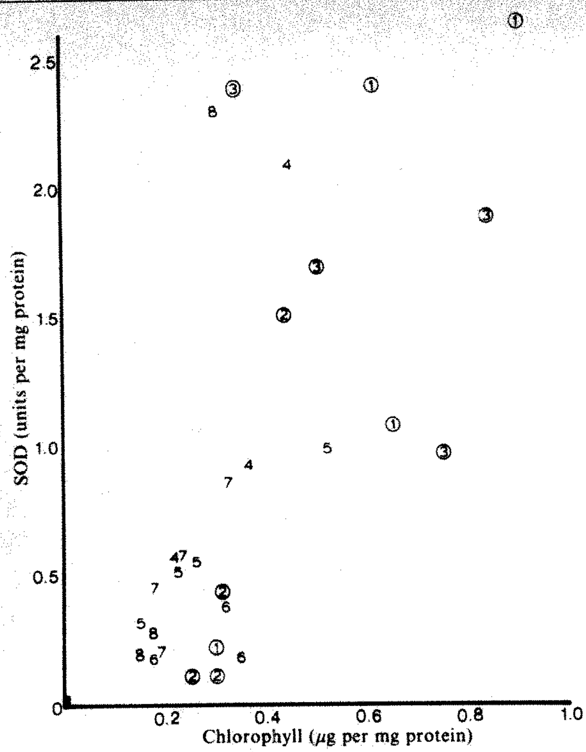


Fig. 2 Correlation between SOD activity and total chlorophyll in individuals from eight different clones of the sea anemone *A. elegantissima*. Clone-mates share numbers and symbiotic anemones collected from the upper areas of the intertidal zone are circled. The mean SOD activity of six aposymbiotic anemones totally lacking chlorophyll is represented by the single solid square (mean SOD activity $= 1.14 \times 10^{-2}$ U per mg animal protein). Regression: $y = 2.575x - 0.013$, $r = 0.738$, $P < 0.001$, $n = 35$. SOD activity, expressed in Units according to McCord and Fridovich[27], was determined using a Clark-type polarographic oxygen electrode and a procedure modified from Tyler[28], and from Marshall and Worsfold[29]. After homogenizing the anemone, the algal cells were removed by centrifugation (3,000g, 15 min) and the SOD activity reported is that of the resulting animal supernatant. Chlorophylls *a* and c_2 were calculated using the equations of Jeffrey and Humphrey[30], following two 10 h extractions in 100% acetone. No chlorophyll could be detected in the homogenization supernatant using this spectrophotometric procedure or by using the more sensitive fluorometric technique of Yentsch and Menzell[31], thus verifying that the algal cells had not been disrupted by the homogenization procedure and that the SOD activity is that of the animal tissue. Protein was determined according to Itzhaki and Gill[32], and, after tissue digestion in 5% KOH, by the technique of Bradford[33].

chemical inhibition or darkness decreased host SOD activity. It therefore appears that host SOD activity is altered in response to the amount of O_2 generated photosynthetically by its intracellular algal symbionts.

The net production of oxygen by alga–invertebrate associations has been interpreted almost exclusively as an indication of the nutritional contribution of the algae to the host[11,12,23], although intracellularly produced oxygen has been implicated as a buffer against environmental hypoxia and as a factor modifying behaviour in sea anemones[24–26]. The high levels of photosynthetically generated oxygen are a potentially deleterious factor in these symbioses. To exploit the benefits of the association while avoiding oxygen toxicity, the host sea anemones must maintain levels of SOD in direct proportion to their algal complement.

We thank Dr Bruce Sidell for technical advice, discussions and critical reading of the manuscript, Dr Leonard Muscatine for comments on the manuscript, and William Zamer for collecting and maintaining the anemones and for helpful discussions. This research was supported by the University of Maine Faculty Research Fund and by NSF grant PCM 79-11027 (Regulatory Biology) to J.M.S.

Received 18 March; accepted 20 April 1982.

1. Halliwell, B. *Cell Biol. int. Rep.* **2,** 113–128 (1978).
2. Fridovich I. *Science* **27,** 462–466 (1977).
3. Yagi, K. & Ohishi, N. in *Biochemical and Medical Aspects of Active Oxygen* (eds Hayaishi, O. & Asada, K.) 299–307 (University of Tokyo Press, Tokyo, 1977).
4. Barthelemy, L., Beland, A. & Chastel, C. *Resp. Physiol.* **44,** 261–268 (1978).
5. Pearse, V. B. *Biol. Bull.* **147,** 641–651 (1974).
6. Wethey, D. S. & Porter, J. W. in *Coelenterate Ecology and Behavior* (ed. Mackie, G. O.) 59–66 (Plenum, New York, 1976).
7. Crossland, C. J. & Barnes, D. J. *Mar. Biol.* **40,** 185–194 (1977).
8. Svoboda, A. & Porrmann, T. in *Nutrition in the Lower Metazoa* (eds Smith, D. C. & Tiffon, Y.) 87–99 (Pergamon, New York, 1980).
9. Trench, R. K., Wethey, D. S. & Porter, J. W. *Biol. Bull.* **161,** 180–198 (1981).
10. Mangum, C. P. & Johansen, K. *Pacific Sci.* (in the press).
11. Trench, R. K. *A. Rev. Pl. Physiol.* **30,** 485–531 (1979).
12. Muscatine, L. in *Primary Productivity in the Sea* (ed. Falkowski, P. G.) 381–402 (Plenum, New York, 1980).
13. Hill, H. A. O. *Ciba Fdn Symp.* **65,** 5–17 (1979).
14. Halliwell, B. in *Superoxide and Superoxide Dismutases* (eds Michelson, A. M., McCord, J. M. & Fridovich, I.) 335–349 (Academic, New York, 1977).
15. Fee, J. A. in *Biological and Clinical Aspects of Superoxide and Superoxide Dismutase* (eds Bannister, W. H. & Bannister, J. V.) 41–48 (Elsevier, Amsterdam, 1980).
16. Fridovich, I. *Science* **201,** 875–880 (1978).
17. Roos, D. & Weening, R. S. *Ciba Fdn Symp.* **65,** 225–262, (1979).
18. Hassan, H. M. & Fridovich, I. in *Enzymic Basis of Detoxication* (ed. Jakoby, W. B.) 311–332 (Academic, New York, 1980).
19. Trench, R. K. *Proc. R. Soc.* **B177,** 237–250 (1971).
20. Fitt, W. K. & Pardy, R. L. *Mar. Biol.* **61,** 199–205 (1981).
21. Asada, K., Kanematsu, S. & Uchida, K. *Archs Biochem. Biophys.* **179,** 243–256 (1977).
22. Steele, R. D. *J. Zool. Lond.* **179,** 387–405 (1976).
23. Muscatine, L., McCloskey, L. R. & Morian, R. E. *Limnol. Oceanogr.* **26,** 601–611 (1981).
24. Pearse, V. B. *Biol. Bull.* **147,** 630–640 (1974).
25. Fredericks, C. A. *Mar. Biol.* **38,** 25–28 (1976).
26. Shick, J. M. & Brown, W. I. *J. exp. Zool.* **201,** 149–155 (1977).
27. Fridovich, I. & McCord, J. *J. biol. Chem.* **244,** 6049–6055 (1969).
28. Tyler, D. D. *Biochem. J.* **147,** 493–504 (1975).
29. Marshall, M. J. & Worsfold, M. *Analyt. Biochem.* **86,** 561–573 (1978).
30. Jeffrey, S. W. & Humphrey, G. H. *Biochem. Physiol. Pflanzen.* **167,** 191–194 (1975).
31. Yentsch, C. S. & Menzel, D. W. *Deep-Sea Res.* **10,** 221–231 (1963).
32. Itzhaki, R. F. & Gill, D. M. *Analyt. Biochem.* **9,** 401–410 (1964).
33. Bradford, M. M. *Analyt. Biochem.* **72,** 248–254 (1976).

Central connections of the retinal ON and OFF pathways

Peter H. Schiller

Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Slaughter and Miller have recently demonstrated that in the isolated eye cup of the mudpuppy and the rabbit, DL-2-amino-4-phosphonobutyric acid (APB) reversibly blocks the ON responses in the retina: on infusion of APB, ON bipolar, ON amacrine and ON ganglion cells become unresponsive; receptors, horizontal cells, OFF bipolars, OFF amacrines and OFF ganglion cells are unaffected[1]. The centre-surround organization of OFF-centre ganglion cells appears unaltered. I have used these findings to determine how the ON-centre and OFF-centre retinal ganglion cells, whose characteristics have been extensively studied in mammals[2–7], exert their influence on the neurones of the central visual system. Recordings made in the lateral geniculate nucleus (LGN) and striate cortex of the rhesus monkey while the retinal ON responses were reversibly blocked with APB showed that in the LGN, retinal APB infusion blocked the responses of ON-centre cells and had little effect on OFF-centre cells. The centre-surround organization of OFF-centre cells was unaffected. APB infusion eliminated the light-edge responses of cortical cells, revealed a dark-edge response in some, but had no discernible effect on orientation and direction specificities. These results suggest that the ON and OFF systems do not interact significantly at the level of the LGN, but do so in the striate cortex.

Surgery was performed on rhesus monkeys under sodium pentobarbital anaesthesia. During recording, animals were paralysed with Flaxedil and were respirated with a 75–25% mixture of NO_2–O_2. The eye was perfused through a tube inserted behind the limbus into the vitreous chamber. The perfusate left the eye through a second tube. The control perfusate, which was similar to that reported by Cunningham and Miller[8], was dripped at the rate of 2–4 ml min^{-1}. A stopcock

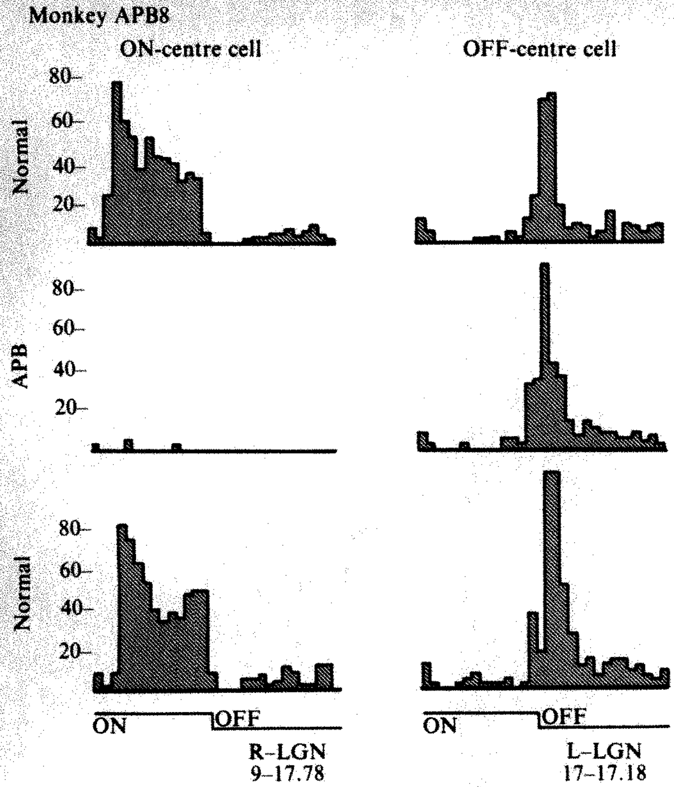


Fig. 1 The effect of retinal APB infusion into the left eye (200 μM APB for 1 min) on two simultaneously recorded single cells, one in layer 6 of the right LGN and the other in layer 2 of the left LGN. The receptive field centre of each cell was stimulated with a small spot of light, using two light beams, presented at the rate of 500 ms on and 500 ms off. Each histogram is based on responses elicited with 30 repeated stimulus presentations. The histograms on top were obtained before APB infusion, those in the middle after the infusion had blocked the retinal on responses, and those on the bottom after the ON-centre cell had recovered. APB abolished the responses of the ON-centre cell and had no significant effect on the OFF-centre cell.

allowed us selectively to infuse a solution which in addition to the perfusate contained 100–800 μM of APB. Electrodes were placed in the LGN and striate cortex of 10 animals. In most cases we recorded simultaneously with two electrodes, either with one in each LGN, or with one in the LGN and the other in the striate cortex. Stationary, flashing spots or moving light bars were presented on a tangent screen with a computer-driven optic bench[9]. The effects of APB infusion were studied at 124 LGN (78 single cells) and 55 cortex sites (41 single cells); their receptive fields were within 12° of the fovea. While recording in the LGN it was often also possible to monitor the activity of retinal ganglion cell axons[10].

In the LGN, brief 1–3-min infusion of 100–200 μM APB effectively blocked the activity of ON-centre cells within 1–3 min. In successful preparations, parvocellular ON-centre cells, most of which have colour-opponent receptive field properties[11–13], were blocked the whole time. Magnocellular ON-centre cells, which have broad-band receptive field properties[11–13], were more resistant to APB infusion; their responses were generally blocked several seconds later than those of parvocellular cells, and in some animals the response was reduced but not entirely eliminated. Therefore, to be certain about the extent to which ON responses were blocked while we studied ON–OFF interactions, we always assessed the responses of cells in the magnocellular lamina. The data reported here are based on those cases where we established that both the magnocellular and parvocellular responses were blocked by our infusions. For the most part, OFF-centre cells were not significantly affected by the APB. Figure 1 shows the effects of APB infusion on an ON-centre and an OFF-centre cell recorded simultaneously with two electrodes in the LGN. The infusion abolished the responses of the ON-centre cell and had little effect on the OFF-centre cell. Recordings from retinal ganglion cell axons showed similar effects.

To determine the extent to which the centre-surround antagonism observed in LGN cells is produced by the convergence of the ON and OFF channels, we selectively stimulated the centre and surround of LGN cells before and during APB infusion. Figure 2 shows one of the several methods we used. The receptive field centre was stimulated with a spot of light

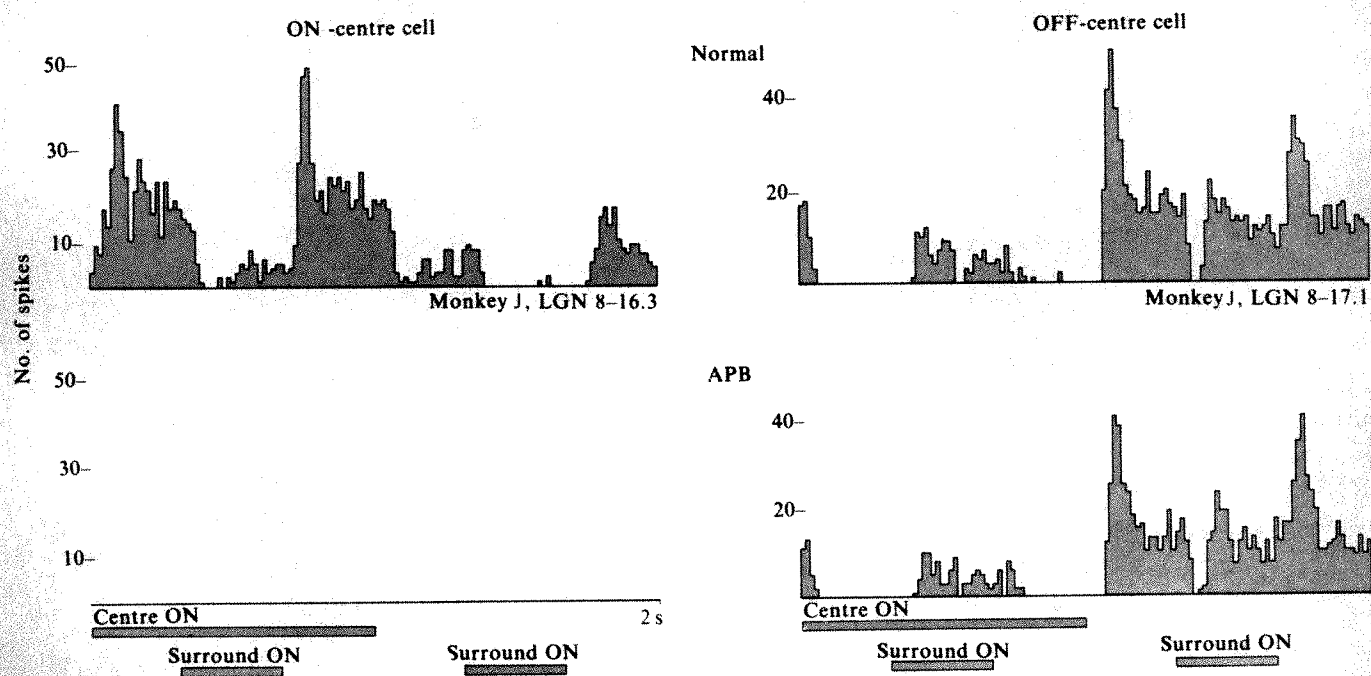


Fig. 2 The responses of an ON-centre and an OFF-centre colour-opponent cell to centre and surround stimulation before and after 200 μM infusion of APB into the eye. The centre stimulus (0.75° green spot for the ON-centre cell and 0.75° red spot for the OFF-centre cell) was flashed for 1 s every 2 s. The surround stimulus (9° red spot for the ON-centre cell and 9° green spot for the OFF-centre cell) was flashed on for 360 ms twice during each cycle, first 320 ms after the centre spot appeared and second 320 ms after the centre spot was extinguished. When the surround stimulus appeared alone, it yielded a response opposite in sign to that produced by the centre stimulus. Each histogram is based on 30 cycles. APB infusion blocked all the responses elicited from the ON-centre cell and had no significant effect on the responses of the OFF-centre cell.

flashed on for 1 s every 2 s. The surround was stimulated with a large spot of a different wavelength, which appeared once with the centre spot on and once with the spot extinguished. Figure 2 shows the responses of an ON- and an OFF-centre cell for such conditions of stimulation. Both were colour-opponent cells, and were stimulated accordingly. The centre of the first cell was activated with a small green spot, while the surround was stimulated with a large red spot. The opposite arrangement was used for the OFF-centre cell, which for its centre required a red stimulus and for its surround a large green spot.

In the ON-centre cell, the surround stimulus, when paired with the centre spot, caused inhibition; when it was presented alone, it yielded a response opposite in sign to that produced by the centre stimulus. APB infusion (200 μM for 1 min) eliminated all the responses of the ON-centre cell, including the OFF response produced by surround stimulation. Similar infusion did not have a differential effect on the centre and surround responses of the OFF-centre cell; the ON response produced by the surround stimulation while the centre spot was on persisted. The centre-surround organization of OFF-centre broad-band cells was also unaffected by APB infusion. These findings suggest that the enhanced centre-surround antagonism of LGN cells, evident in both cat and monkey[11,12,14], is not produced, in the monkey at least, by interactions between the ON and OFF channels, as some investigators have suggested for the cat[14,15].

In the striate cortex we examined the effects of APB infusion on the light- and dark-edge responses of cortical cells, their directionality and their orientation specificity[9,16]. When the ON responses were blocked by APB infusion in both the parvo-

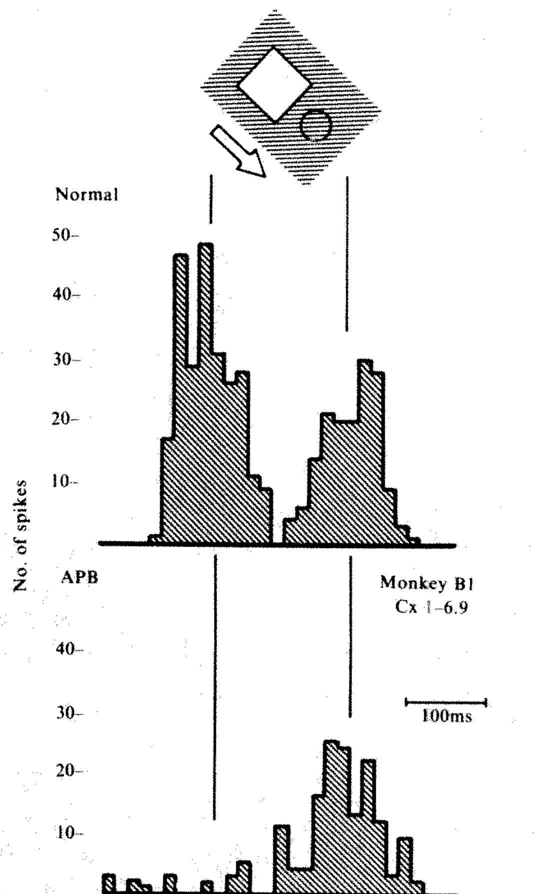


Fig. 3 The responses of a complex cell in striate cortex to a moving 1° wide light bar, as shown in the inset, before and after the infusion of 200 μM APB into the eye. The stimulus was moved across the receptive field at 4° per s. Data were collected for 30 repeated trials. The vertical lines show expected location of the light-edge (L) and dark-edge (D) responses. APB blocked the light-edge response and had little effect on the dark-edge response.

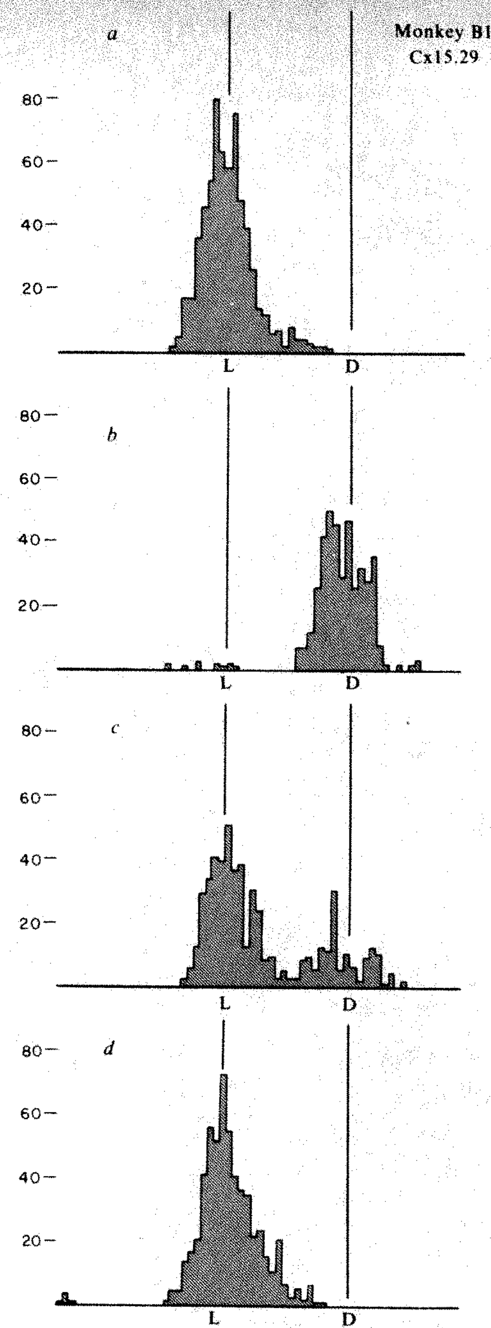


Fig. 4 The responses of a simple cell in the striate cortex before and after the infusion of APB. A 1° wide light bar was swept across the receptive field from left to right at 2° per s. Each histogram is the result of 30 repeated sweeps. The vertical lines show the expected location of the light-edge (L) and dark-edge (D) responses. *a*, Normal response during infusion of control perfusate; *b*, response after infusion of 100 μM APB had taken effect; *c*, partial recovery after termination of APB infusion; *d*, full recovery following termination of APB infusion. APB infusion uncovered a dark-edge response for this cell.

cellular and magnocellular layers of the LGN, the light-edge response of cortical cells was eliminated. An example obtained from a complex cell is shown in Fig. 3. The cell was stimulated with a moving light bar, which elicited a response to both the leading light edge and the trailing dark edge. Infusion of 200 μM APB eliminated the light-edge response, suggesting that it is produced by the input from the ON channel. In some cortical cells more complicated interactions were evident. Figure 4 shows one such example. This simple cell[9], when stimulated in the same fashion as the cell shown in Fig. 3, responded only to a light edge. Infusion of 100 μM of APB eliminated the light-edge response, and, unexpectedly, uncovered a dark-edge response. Part of the way to recovery, both responses were

evident. Recordings in magnocellular LGN have shown that APB infusion fully blocked the ON responses. These observations suggest that in addition to convergent excitation, significant inhibitory interactions can also occur between the ON and OFF channels in the striate cortex.

The orientation and direction specificities of cortical cells were assessed by examining the responses of cells to light and dark edges which were moved across receptive fields in various orientations in random order before, during and after blocking the retinal ON system. The directionality and orientation selectivity of the cortical cells we studied were unaltered by APB infusion, suggesting that these attributes are not produced by interactions between the ON and OFF channels[17]. Work carried out concurrently by J. C. Horton and H. Sherk using a different procedure has yielded similar results in the cat[18].

Thus, my results suggest that in the rhesus monkey ON-centre and OFF-centre retinal ganglion cells drive predominantly ON-centre and OFF-centre LGN cells respectively. The surround inhibition observed in LGN cells does not appear to be a product of the interaction between the ON and OFF channels. In the striate cortex the ON and OFF channels converge, with the light-edge response often produced by the input from ON-centre cells and the dark-edge response produced by OFF-centre cells. Inhibitory interactions between these channels are evident in the cortex. The orientation and direction specificity of cortical cells were unaffected by blocking the ON system, indicating that each channel has access to the mechanisms responsible for these attributes.

I thank Denice Couch-Helwig, Martha Rhodes and Nancy Stec for their assistance. This work was supported in part by NIH grants EY 02621 and EY 00676 and NSF grant BNS 8019714.

Received 13 October 1981; accepted 31 March 1982.

1. Slaughter, M. M. & Miller, R. F. *Science* **211**, 183 (1981).
2. Cleland B. G. & Levick, W. R. *J. Physiol., Lond.* **240**, 421 (1974).
3. DeMonasterio F. M. & Gouras, P. *J. Physiol., Lond.* **251**, 167 (1975).
4. Enroth-Cugell, C. & Robson, J. G. *J. Physiol., Lond.* **187**, 517 (1966).
5. Kuffler, S. W. *J. Neurophysiol.* **16**, 37 (1953).
6. Schiller, P. H. & Malpeli, J. G. *J. Neurophysiol.* **40**, 428 (1977).
7. Stone, J. & Fukuda, Y. *J. Neurophysiol.* **37**, 722 (1974).
8. Cunningham, R. & Miller, R. F. *Brain Res.* **117**, 341 (1976).
9. Schiller, P. H., Finlay, B. L. & Volman, S. F. *J. Neurophysiol.* **39**, 1288 (1976).
10. Schiller, P. H. & Malpeli, J. G. *Brain Res.* **137**, 387 (1977).
11. Dreher, B., Fukuda, Y. & Rodieck, R. W. *J. Physiol., Lond.* **258**, 433 (1976).
12. Schiller, P. H. & Malpeli, J. G. *J. Neurophysiol.* **41**, 788 (1978).
13. Wiesel, T. N. & Hubel, D. H. *J. Neurophysiol.* **29**, 1115 (1966).
14. Hubel D. H. & Wiesel, T. N. *J. Physiol., Lond.* **155**, 383 (1961).
15. Maffei, L. & Fiorentini, A. *J. Neurophysiol.* **35**, 65 (1972).
16. Hubel, D. H. & Wiesel, T. N. *J. Physiol, Lond.* **195**, 215 (1968).
17. Hubel, D. H. *J. opt. Soc. Am.* **53**, 58 (1963).
18. Horton, J. C. *Neurosc. Abstr.* **7**, 11.8 (1981).

Electrophysiological action of kainic acid and folates in the *in vitro* olfactory cortex slice

C. R. Auker*, D. J. Braitman† & S. L. Rubinstein

Physiology Department, Armed Forces Radiobiology Research Institute, Bethesda, Maryland 20814, USA

Evidence has mounted that kainic acid (KA), a powerful neuroexcitatory and neurotoxic agent, acts at a specific class of receptors distinct from those mediating the excitatory actions of glutamate[1–3]. This has prompted a search for the endogenous ligand for the KA-specific receptors. Recently, Ruck *et al.*[4] reported that the naturally occurring folic acid derivative methyltetrahydrofolate (MTHF) is a potent and specific competitor for KA-binding sites in rat cerebellar membranes, having one-tenth the binding activity of KA at these sites. They suggested that MTHF may be the endogenous KA receptor ligand. Olney *et al.*[5] found that injections of various folates—MTHF, folic acid (pteroyl-L-glutamic acid, PGA), or folinic acid (formyltetrahydrofolate, FTHF)—into rat amygdala and striatum produced KA-like seizures and associated brain damage at sites distant from the site of injection. However, these folates did not mimic the direct neurotoxic effects of KA at the injection site. Furthermore, Roberts *et al.*[6] reported that MTHF injected into rat cerebellum produced slower and less complete neuronal degeneration than did KA and, unlike KA and glutamate, did not increase cerebellar cyclic GMP levels. In light of the contradictory results from these binding, histological and biochemical studies, we have tested the electrophysiological equivalency of KA and the folates in rat olfactory cortex slices. KA is a potent agonist at the receptors of the terminal synapses of the lateral olfactory tract (LOT); in fact, the natural transmitter at this synapse appears to act at a KA-preferring receptor[7]. Also, olfactory cortex is extremely sensitive to the neurotoxic action of KA[8]. We report here that MTHF, FTHF and PGA had little or no effect on LOT-stimulated field potentials even at millimolar concentrations, whereas KA exhibited excitatory effects at concentrations as low as 5×10^{-7} M. It therefore seems unlikely that MTHF or the other folates are endogenous ligands for the KA receptor in olfactory cortex.

* Present address: Ophthalmology Department, Georgetown University Medical Center, Reservoir Road, Washington, DC, USA.
† To whom correspondence should be addressed.

Thin slices (300–450 μm) of rat olfactory cortex were cut and incubated as previously described[7]. Each slice was transferred to a recording chamber that was constantly perfused with oxygenated normal Ringer's solution. The LOT was stimulated with double shocks of 20–50 V intensity and 50 μs duration via a tungsten bipolar electrode placed across the tract. Paired

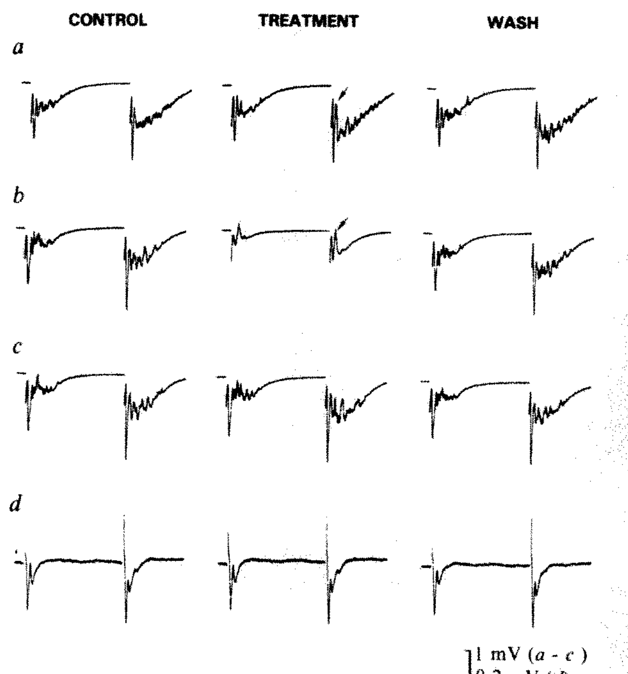


Fig. 1 Effects of KA and folates on LOT-stimulated field potentials in rat olfactory cortex slices. Each trace shows two field potentials in response to double shock stimulation of LOT. As previously reported, the second field potential of each pair is larger than the first[7,10]. Initial downward deflection in each trace is the beginning of the e.p.s.p.[11]. Rapid upward deflection is due to the superimposition of the population spike onto the slow wave component of the field potential[10,11]. The arrows in *a* and *b* indicate the population spike in the second field potential. Spiking in the late wave of the field potential indicates asynchronous neural firing. Slices were superfused with: *a*, 5×10^{-7} M KA for 15 min; *b*, 10^{-5} M KA for 5 min; *c*, 10^{-4} M FTHF for 15 min; and *d*, PGA for 20 min. The duration of wash was equal to or greater than duration of treatment. FTHF was prepared in oxygenated Ringer's just before use and was protected from light during experiment.

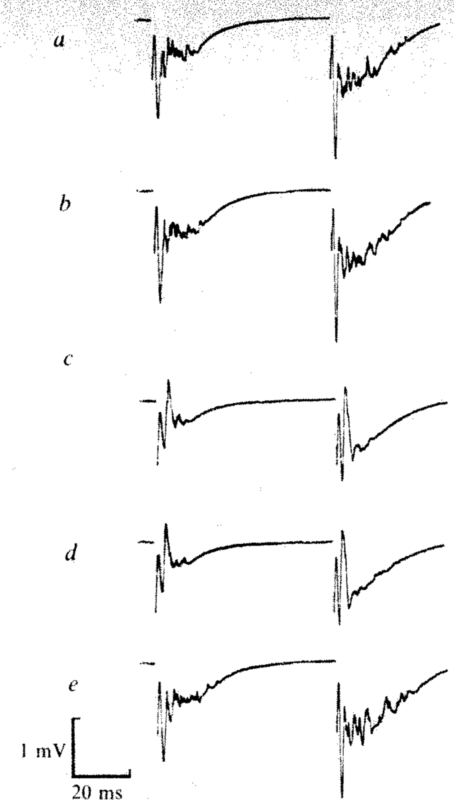


Fig. 2 Lack of agonist and antagonist activity of MTHF in rat olfactory cortex slice. *a*, Control: cortical field potentials evoked by double shock to LOT. Effect of treatment *b*, for 10 min with 10^{-4} M MTHF; *c*, for 8 min with 5×10^{-6} M KA. *d*, Effects of combined treatment for 7 min and 30 s with both 10^{-4} M MTHF and 5×10^{-6} M KA. *e*, Wash for 31 min. MTHF was prepared in oxygenated Ringer's just before use and was protected from light during experiment.

shocks were separated by 60 ms and administered every 4 s. Shock intensities were adjusted to produce field potentials of maximum amplitude. Glass micropipettes with ~50 μm cut tips were filled with normal Ringer's solution and used to record field potentials from the pial surface of the slice. After recording stable control field potentials for several minutes, the perfusate was changed to a Ringer's solution containing either KA, PGA, MTHF, FTHF, or a combination of KA and MTHF. Concentrations of these agents ranged from 10^{-8} to 10^{-3} M, and perfusion times from 5 to 30 min. This was followed by a wash in normal Ringer's for a period of time equal to or greater than that of the drug treatment.

KA administered in concentrations $>10^{-8}$ M exhibited dose-dependent excitatory effects on the LOT-stimulated field potential. The lowest effective bath concentration, 5×10^{-7} M KA, increased the amplitude of the population spike, indicating an increase in synchronous neuronal firing, but had little effect on the slow wave representing the population excitatory post-synaptic potential (e.p.s.p.) (Fig. 1*a*). At higher concentrations (Figs 1*b*, 2*c*), KA application resulted in an initial increase in the width and amplitude of the population spike followed by a decrease in amplitude of the e.p.s.p. As the e.p.s.p. approached the point of abolishment, the population spike increased to maximum size and then decreased and disappeared. These events appear to represent excitation followed by depolarization block. The actions of KA were reversible if wash with normal Ringer's was initiated no later than the time of disappearance of the population spike. In contrast to KA, PGA (Fig. 1*d*), FTHF (Fig. 1*c*) and MTHF (Fig. 2*b*) had little or no effect on the LOT-stimulated e.p.s.p. or population spike even when administered in concentrations as high as 1 mM. At concentrations $\geq10^{-4}$ M, these substances occasionally increased asynchronous spiking during the late stage of the field potential, an action consistent with the weak excitatory effects demonstrated for PGA and FTHF in cat cortex[9]. As none of the folates reproduced the effects of KA (even at 2,000 times the KA concentration), it seems unlikely that the folates are physiological ligands for the KA receptor.

As MTHF has been reported to be a potent competitor for KA-binding sites in rat cerebellum[6] but does not seem to be an agonist in rat olfactory cortex, we tested whether MTHF is an antagonist at the KA receptors. MTHF (10^{-4} M) and KA (5×10^{-6} M) were individually and then simultaneously bath-applied to the slice during LOT stimulation. KA alone decreased the amplitude of the e.p.s.p. and increased the amplitude of the population spike, whereas MTHF alone had no effect; in combination with KA, MTHF neither diminished nor slowed the effects of KA on the e.p.s.p. or the population spike (Fig. 2). Thus MTHF does not appear to be an antagonist at the KA receptor in olfactory cortex.

Evidence exists that the folates can excite central nervous system neurones. Davies and Watkins[9] showed that cat cortical neurones were weakly excited by PGA and FTHF, and Loots *et al.* (cited in ref. 4) have shown that MTHF depolarizes frog spinal cord neurones. However, neither of these studies examined the relationship between the actions of the folates and KA. We have directly compared the electrophysiological effects of the folates and KA in a system that has been demonstrated to have KA receptors, and in which the endogenous transmitter has KA-like activity[7]. In this system, neither the folate derivatives (FTHF and MTHF) nor folic acid itself reproduced the effects of KA. Thus the folates cannot be considered to be the natural KA ligand at the terminal synapses of the LOT. The apparent discrepancy between our electrophysiological data and the binding data of Ruck *et al.*[4] may reflect (1) existence of different KA receptors in different brain regions, or (2) coexistence, in the same brain region, of at least two types of KA receptor, one that binds KA but not the folates and has electrophysiological effects, and another that binds KA and the folates but has no electrophysiological effects measurable by our techniques.

This work was supported by the Armed Forces Radiobiology Research Institute, Defense Nuclear Agency, under research work unit MJ 00005. The views presented in this paper are those of the authors. No endorsement by the Defense Nuclear Agency has been given or should be inferred.

Received 18 January; accepted 16 March 1982.

1. Hall, J. G., Hicks, T. D. & McLennan, H. *Neurosci. Lett.* **8**, 171–175 (1978).
2. Baudry, M. & Lynch, G. *Eur. J. Pharmac.* **57**, 283–285 (1979).
3. Watkins, J. C. *Adv. Biochem. Psychopharmac.* **29**, 205–212 (1981).
4. Ruck, A., Kramer, S., Metz, J. & Brennan, M. J. W. *Nature* **287**, 852–853 (1980).
5. Olney, J. W., Fuller, T. A. & de Gubareff, T. *Nature* **292**, 165–166 (1981).
6. Roberts, P. J., Foster, G. A. & Thomas, E. M. *Nature* **293**, 654–655 (1981).
7. Hori, N., Auker, C. R., Braitman, D. J. & Carpenter, D. O. *Cell. molec. Neurobiol.* **1**, 115–120 (1981).
8. Olney, J. W. & de Gubareff, T. in *Kainic Acid as a Tool in Neurobiology* (eds McGeer, E., Olney, J. W. & McGeer, P. L.) 201–217 (Raven, New York, 1978).
9. Davies, J. & Watkins, J. C. *Biochem. Pharmac.* **22**, 1667–1668 (1973).
10. Richards, C. D. *J. Physiol., Lond.* **222**, 209–231 (1972).
11. Richards, C. D. & Sercombe, R. *J. Physiol., Lond.* **211**, 571–584 (1970).

Neurone differentiation in cell lineage mutants of *Caenorhabditis elegans*

J. G. White*, H. R. Horvitz† & J. E. Sulston*

* MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK
† Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

The nematode *Caenorhabditis elegans* develops by an essentially invariant sequence of cell divisions[1–3] leading to an adult complement of 959 somatic cells. In this organism cell fate is correlated with cell lineage, suggesting that genealogy may be a determining factor for the differentiated state of a cell. The study of mutants with altered cell lineages may help elucidate the precise mechanisms by which cell fate is decided. Several

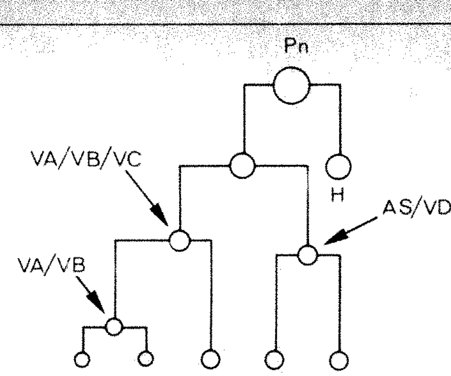
0028-0836/82/240584—04$01.00

cell lineage mutants have been isolated and characterized[4,5], some having more and some fewer cell divisions than wild type. We have now investigated the cell types produced by two cell lineage mutants; these mutants exhibit blocks in certain terminal or near terminal cell divisions, which in normal animals generally give rise to daughter cells that differentiate into distinctly different cell types. We find that the blocked cells in the mutants generally exhibit the differentiated characteristics of only one of the two daughter cells that normally would be produced. The differentiated state of the blocked precursors may be due to an intrinsic dominance of one cell type over another in what is essentially a fused cell, and/or it may reveal the state of commitment of the precursor in wild-type animals.

We have studied two very similar mutants *unc*-59(*e*1005) and *unc*-85(*e*1414), both of which are variably blocked in some of the later divisions of the lineages that produce the adult complement of motoneurones in the ventral nerve cord[4,5]. These lineages have been characterized in wild-type animals by following the development of living animals in the light microscope[1]. The nuclei of 12 precursor cells (designated P1–P12) migrate into the ventral cord during the first larval stage; each cell then divides to produce a neuroblast and a hypodermal cell. The 12 neuroblasts (together with an additional neuroblast present at hatching) undergo identical sequences of divisions (Fig. 1), each producing 5 cells that intercalate with the pre-existing juvenile motoneurones to form a single row of cells along the ventral cord. All cleavages have longitudinal spindle axes, and descendant cells maintain their relative antero-posterior positions throughout development. Certain cells derived from some of the neuroblasts die soon after their formation; the pattern of cell death is invariant.

The structure of the ventral cord of wild-type animals has been deduced by reconstructions from electron micrographs of serial sections[6]. In neuroblast lineages that do not contain cell deaths, the five cells produced differentiate into five distinct classes of motoneurone designated VA, VB, VC, AS and VD (Fig. 1). Classes are defined on the basis of patterns of synaptic connections (see Pn, Table 2) and morphology (Fig. 2). Either criterion is sufficient to assign a cell to a class.

The post-embryonic development of ventral cord cells was observed in several *unc-59* and *unc-85* animals. In these mutants, P-cell-derived neuroblasts produced fewer progeny cells than normal (Fig. 1), because some cell divisions failed. The set of affected cells varied between animals. Failures resulted in polyploid cells; sometimes the nuclei failed to divide after DNA replication, and sometimes apparently normal nuclei formed but then fused to form a tetraploid nucleus or remained separated in a binucleate cell[5]. *unc-59* and *unc-85* individuals that displayed several examples of undivided ventral cord nuclei were selected for reconstruction from electron micrographs of serial sections.



VA	VB	VC	AS	VD	P3 P4 P5 P6 P7 & P8
VA	VB	X	AS	VD	P2 P9 & P10
VA	X	X	AS	VD	P11
AVF	VB	X	AS	VD	P1
VA	X	X	PDB	VD	P12

Fig. 1 The ventral cord precursor cells P1–P12 are hypodermal cells (J.G.W., unpublished observations). Their nuclei migrate into the ventral cord and each divides to produce a ventral hypodermal cell (H) (which functionally replaces the mother) and a neuroblast. All 12 neuroblasts then undergo identical series of divisions to produce five descendants[1]. All spindle axes are longitudinal; in the diagram anterior daughters are drawn to the left and posterior daughters to the right. The fates of the descendants from each precursor are shown. Cell types have been assigned on the basis of cell morphology and synaptic connectivity[6]. Precursors P3–P8 produce five classes of ventral cord motoneurone (VA, VB, VC, AS, VD). In the P1–P2 and P9–P12 lineages, cells lineally equivalent to VC neurones undergo programmed cell death (X), as do cells lineally equivalent to VB neurones derived from P11–P12. Alternative fates are seen for some of the cells derived from the two precursors at the ends of the cord (P1, P12): P1 produces an AVF interneurone instead of a VA, and P12 produces a PDB interneurone instead of an AS. (AVF and PDB are two distinct classes of interneurone which are quite dissimilar to the motoneurone classes.) These alternative fates are probably specified by local interactions[1,14]. In the *unc-59* and *unc-85* animals, certain cells which divide in the wild-type fail to do so. Such cells are labelled according to the fate of their normal descendants; for example, an AS/VD cell should have divided to produce an AS and a VD neurone.

In both *unc-59* and *unc-85* animals all blocked cells were found to have differentiated into neurones. Most of these neurones could be unequivocally identified as belonging to one of the motoneurone classes normally produced from these lineages; specifically, the blocked cells acquired the differentiated characteristics of one of the daughter cells that would normally be produced. These results are summarized in Table 1. All blocked VD/AS precursors differentiated into cells that had all the distinguishing characteristics of VD motoneurones (Fig. 2, Tables 1, 2). VA/VB or VA/VB/VC precursors differentiated into VB-like cells in the anterior ventral cord and into VA-like cells in the posterior ventral cord (Tables 1, 2). The choice of VA versus VB for the blocked precursors may be related to the regional preference exhibited by the progeny of these precursors in wild-type development: a VB but no VA is produced by the anterior P1 lineage, whereas a VA but no VB is produced by the posterior P11 and P12 lineages (Fig. 1).

Several differences were apparent between the blocked precursors and the corresponding cell types in wild-type animals: the blocked cells had more synaptic inputs (Table 1), larger nuclei and extra branches (Fig. 2). These extra branches were generally devoid of synapses and had none of the characteristics of the other cell type. Even more extensive branching has been seen from precursors that are blocked earlier in ventral cord lineages in the mutant *lin-5* (ref. 7). The supernumerary branches may be a consequence of the polyploid nature of these cells; perhaps the total process length produced from a given neurone is proportional to its ploidy. The embryonically produced neurones in this region (that is, the DA, DB and DD motoneurones[6,8]) were normal in morphology and connectivity.

The observation that characteristics of some cell types were expressed normally in the blocked cells implies that the normal

Table 1 Post-embryonic development of P cells from *unc-85* and *unc-59* animals

Mutant	Blocked precursor		Cell type
unc-85	P3	VA/VB/*	VB
	P3	AS/VD	VD
	P4	VA/VB/*	VB
	P5	VA/VB	VB
	P6	VA/VB/*	VB‡
	P6	AS/VD	VD
unc-59			
Animal 1	P3	VA/VB/VC	VB
	P3	AS/VD	VD
Animal 2	P8	VA/VB	VA
	P8	AS/VD	VD
	P9	VA/VB/†	VA
	P9	AS/VD	VD
	P10	VA/VB/†	VA
	P10	AS/VD	VD
	P11	VA/†/†	VA‡
	P11	AS/VD	VD

*In these lineages a VC nucleus was produced but was not present when the adult was reconstructed. These nuclei probably either fused to make a VA/VB/VC or died.

† The cells lineally equivalent to VC in the P1–P2 and P9–P12 lineages and to VB in the P11–P12 lineages normally undergo programmed cell death[1].

‡ Axons failed to grow out of these cells although they had the synaptic contacts characteristic of the cell types indicated.

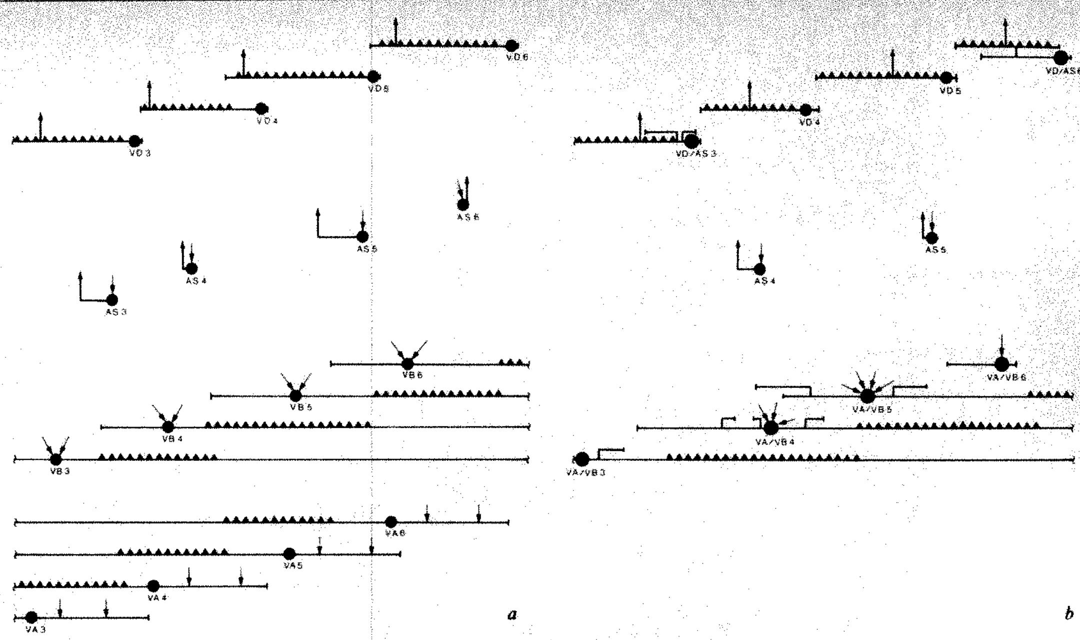


Fig. 2 The morphologies of the motoneurones derived from the precursors P3–P6 are shown for wild type (*a*) and *unc-85* (*b*). The wild-type structures are as described in ref. 6, and the mutant structures were derived, using similar techniques, by reconstruction from 3,000 serial section electron micrographs from an animal of known lineage. Four of the five motoneurone classes derived from the P cells are shown: VA, VB, AS and VD (class VC has been omitted, because it has few distinguishing characteristics). The extent of the processes in the ventral cord and the positions of the cell bodies (-●-), neuromuscular junctions (NMJs) (▲▲▲), commissures (|→) and synaptic input (|←) are shown. The horizontal axis is equivalent to the longitudinal axis of the animal (anterior is drawn to the left); the vertical axis has been used simply to separate the neurones into their respective classes. VA neurones, which have anteriorly directed axons, form a regular sequence of regions of motor activity with little or no overlap between adjacent members of a class. VB neurones are similar but have posteriorly directed axons. VA neurones receive their synaptic input in a dendritic region near the cell body on a branch opposite the axon; VB neurones receive their synaptic input predominantly at the cell body. The axons from class AS motoneurones lead to the dorsal cord via commissures and innervate dorsal muscles. They have a short dendritic region in the ventral cord. The axons from VD neurones end abruptly in gap junctions to axons from adjacent VD neurones. The dendritic regions behave in a similar fashion in the dorsal cord (not shown) and connect to the ventral cord via commissures. In the *unc-85* animal, the VA/VB-blocked precursors from P3–P5 had morphologies similar to wild-type VB neurones, forming NMJs from posteriorly directed axons. The blocked VA/VB precursor from P6 had failed to grow an axon. The AS/VD precursors from P4 and P5 had divided normally, and each produced apparently normal AS and VD daughters. The blocked AS/VD cells from P3 and P6 had the characteristic morphologies of VD motoneurones with their processes forming NMJs and ending abruptly in gap junctions with adjacent normal VD motoneurones. The only differences between the morphologies of these blocked precursors and those of their normal posterior daughters is that the former have larger nuclei, a few extra branches and more synaptic inputs.

complement of wild-type cell divisions is not a necessary prerequisite for the differentiation of these cell types; however, some other cell cycle event (such as DNA replication[15]) may be required. Blast cells that are blocked earlier in a lineage either by drugs[9,10] or in the mutant *lin-5* (ref. 7), also exhibit characteristics of their normal descendants, but in these cases characteristics of several cell types may be present in one polyploid cell. This may be a result of either localized differentiation or excessive dilution of regulatory elements within these abnormally large, highly polyploid cells. The isolation and characterization of mutants that are blocked at intermediate levels may shed some light on these differences.

Not all late divisions failed in these mutants, and those that did not fail produced normal cells. This observation indicates that the suppression of the differentiated characteristics of certain cell types seen in the blocked precursors does not simply

Table 2 Number of synapses formed by blocked precursors and typical synapses of wild-type ventral cord neurones (Pn)[6]

	AVA	AVB	VA	VB	VC	AS	VD	NMJ	
unc-85									
P3 VA/VB/*	‡	‡						15	
P3 AS/VD				§ 1			= 1	19	
P4 VA/VB/*		= 6						20	
P5 VA/VB	= 1	= 9		= 5				8	
P6 VA/VB/*		= 1		= 5					No axon
P6 AS/VD				§ 3			= 1	14	
unc-59 (1)									
P3 VA/VB/VC		= 15					# 1	5	
P3 AS/VD							= 6	14	
unc-59 (2)									
P8 VA/VB	§ 6 = 6							‡	
P8 AS/VD			§ 3				= 2	11	
P9 VA/VB/†	§ 3 = 5						# 3	22	
P9 AS/VD			§ 3				= 2	13	
P10 VA/VB/†	§ 9 = 5						# 3	16	
P10 AS/VD							= 2	10	
P11 VA/†/†	§ 7								No axon
P11 AS/VD							= 2	12	
Wild type									
Pn VA	§ 4 = 3					= 1	# 1	16	
Pn VB		= 2		= 2			# 2	9	
Pn VC					§ 3 = 5		# 7	2	
Pn AS	§ 2 = 1	§ 2	= 1						
Pn VD			§ 1	§ 2	§ 7		= 1	26	

Values for wild type are the mean values from precursors P3 to P5 obtained from a reconstruction of a wild-type animal. #, Presynaptic; §, postsynaptic; =, gap junctions.

* In these lineages a VC nucleus was produced but was not present when the adult was reconstructed. These nuclei probably either fused to make a VA/VB/VC or died.

† The cells lineally equivalent to VC in the P1–P2 and P9–P12 lineages and to VB in the P11–P12 lineages normally undergo programmed cell death[1].

‡ Processes in regions that normally would display these synapses were out of the region of reconstruction.

reflect the inability of the mutants to produce these cell classes but implicates the failure to divide as the cause of their suppression.

The lack of expression of one cell type in the undivided, probably tetraploid precursors of *unc-59* and *unc-85* may be because there is an intrinsic mutual exclusivity in the expression of cell type such as has been described for certain fused cells[11] and cultured sympathetic neurones[12]. An alternative interpretation of these observations is that the differentiated state of the precursors reflects the state of commitment of these cells in wild-type animals; thus, one of the daughters would normally inherit the state of commitment of the mother. Some support for this notion is provided by observations of mutants in the genes *unc-86* and *lin-4*; these mutants undergo extra cell divisions in specific lineages. In these cases one of the daughters reiterates the characteristic divisions of the mother in a stem cell manner[13], indicating that it has the same state of commitment as its mother cell. These two interpretations are not necessarily incompatible if cells may only express one state of commitment at any time because of an intrinsic mutual exclusivity of cell states.

The electron microscopy was done by N. Thomson and Marilyn Anness. We thank S. Brenner and W. Fixsen for helpful discussions. H.R.H. was supported by postdoctoral fellowships from the Muscular Dystrophy Associations of America and the USPHS and by USPHS grants GM24663 and GM24943.

Received 25 January; accepted 28 April 1982.

1. Sulston, J. E. & Horvitz, H. R. *Devl Biol.* **56**, 110–156 (1977).
2. Kimble, J. & Hirsh, D. *Devl Biol.* **70**, 396–417 (1979).
3. Deppe, U. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 376–380 (1978).
4. Horvitz, R. & Sulston, J. *Genetics* **96**, 435–454 (1980).
5. Sulston, J. & Horvitz, R. *Devl Biol.* **82**, 41–55 (1981).
6. White, J., Southgate, E., Thomson, N. & Brenner, S. *Phil. Trans. R. Soc.* B**275**, 327–348 (1976).
7. Albertson, D., Sulston, J. & White, J. *Devl. Biol.* **63**, 165–178.
8. White, J., Albertson, D. & Anness, M. *Nature* **271**, 746–766 (1978).
9. Whittaker, J. R. *Proc. natn. Acad. Sci. U.S.A.* **70**, 2096–2100 (1973).
10. Laufer, J., Bazzicalupo, P. & Wood, W. *Cell* **19**, 569–577 (1980).
11. Davidson, R. L. in *Somatic Cell Hybridization* (eds Davidson, R. L. & de la Cruz, F.) 131–150 (Raven, New York, 1974).
12. Reichardt, L. & Patterson, P. *Nature* **270**, 147–151 (1977).
13. Chalfie, M., Horvitz, R. & Sulston, J. *Cell* **24**, 59–69 (1981).
14. Sulston, J. E. & White, J. G. *Devl Biol.* **78**, 577–598 (1980).
15. Satoh, N. & Susumu, I. *J. Embryol. exp. Morph.* **61**, 1–13 (1981).

Ovarian secretion of oxytocin is stimulated by prostaglandin

A. P. F. Flint & E. L. Sheldrick

ARC Institute of Animal Physiology, Babraham, Cambridge CB2 4AT, UK

Oxytocin induces regression of corpora lutea when administered to several domestic species[1–3] and a role for endogenous oxytocin in controlling the ovarian cycle is supported by the delay in luteal regression observed in sheep which have been immunized against oxytocin[4]. Although the posterior pituitary is the major source of oxytocin released at parturition and before suckling, recent evidence suggests that during the oestrous cycle the ovary may also secrete oxytocin. Circulating oxytocin and progesterone levels change in parallel during the oestrous cycle and after ovariectomy[5–7], and the corpus luteum[8] (but not other ovarian tissues [A.P.F.F. and E.L.S., unpublished]) contains high concentrations of oxytocin. We now extend this evidence by showing that in sheep, the concentration of oxytocin in ovarian venous blood exceeds that in arterial blood, and we find that ovarian secretion of oxytocin is stimulated by a prostaglandin $F_{2\alpha}$ analogue.

Experiments with Estrumate (cloprostenol, ICI 80996) have shown that when this potent prostaglandin $F_{2\alpha}$ analogue is used to cause luteal regression, the resulting decline in plasma progesterone concentration is accompanied by a drop in jugular venous concentrations of oxytocin during the 1–2 days following treatment (A.P.F.F. and E.L.S., unpublished). However, within 10 min of administering Estrumate there is a transient rise in plasma oxytocin which lasts up to 40 min. The finding that this transient increase is absent in ovariectomized sheep prompted us to measure the veno–arterial concentration difference for oxytocin across the ovary.

Anaesthesia was induced in four Clun Forest ewes 11–14 days after oestrus with intravenous pentobarbitone sodium and maintained after intubation with fluothane in oxygen. Mean oestrous cycle length in the flock was 16.5 days. Ovarian venous blood was collected from ovaries containing corpora lutea by inserting a polyvinyl catheter into an ovarian vein within 2 cm of the ovary and passing it towards the ovary. In two animals catheters were also inserted in ipsilateral utero-ovarian veins and directed away from the ovary. To allow collection of arterial blood and blood draining the posterior pituitary catheters were also inserted into a branch of the contralateral uterine artery and into the external jugular vein. Plasma was prepared from heparinized blood samples and oxytocin and progesterone levels measured by specific radioimmunoassays[6]. A high concentration of progesterone in the ovarian and utero-ovarian veins confirmed that blood was collected from an ovary bearing a corpus luteum.

Estrumate (125 μg, intramuscularly) caused a rapid (within 5 min) release of oxytocin from the ovary, which was maximal 10–20 min after treatment (Fig. 1). At peak oxytocin titres a veno–arterial concentration difference of up to 20 ng ml^{-1} was measured across the ovary (in an animal bearing two corpora

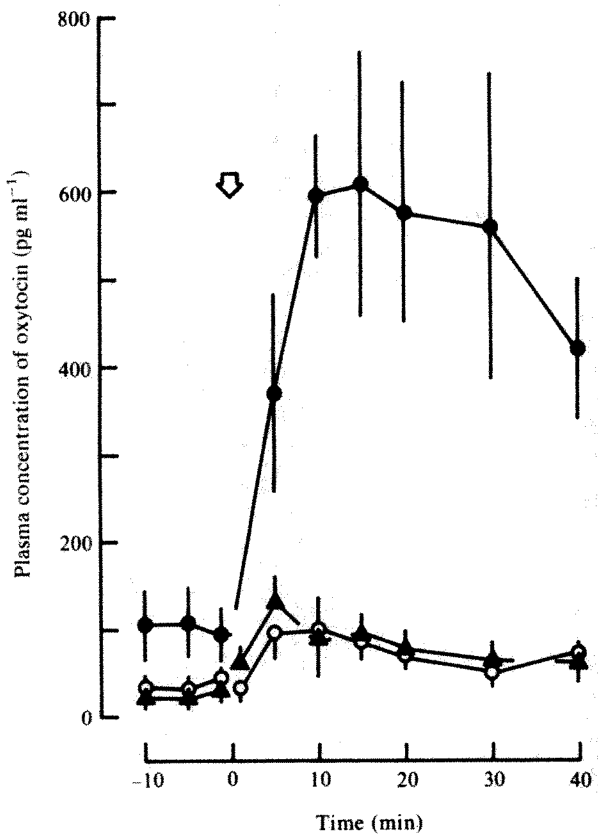


Fig. 1 Oxytocin concentrations measured by radioimmunoassay[6] in plasma from the ovarian vein (●), uterine artery (○) and jugular vein (▲) of four sheep 11–14 days after oestrus. Values are mean ± s.e.m. Ovarian venous concentrations from one animal bearing two corpora lutea in the sampled ovary have been halved; thus values are per corpus luteum. Estrumate (125 μg) was administered intramuscularly at $t = 0$ (arrowed). Anaesthesia was induced with pentobarbitone sodium and maintained with fluothane in oxygen. The antibody used in the radioimmunoassay showed negligible (<0.01%) cross-reactivity with arginine-vasopressin, vasotocin, luteinizing hormone-releasing hormone, ovine luteinizing hormone, progesterone and cloprostenol.

lutea in the sampled ovary; data not included in Fig. 1); ovarian venous concentrations exceeded those in the utero-ovarian vein by a factor of 4.3 (mean of six samples taken simultaneously; $P < 0.005$ by analysis of variance). There was no detectable secretion by the pituitary either before or after administration of Estrumate; the mean concentration ratio jugular vein/artery for oxytocin was 1.1 ($n = 29$ samples). The concentration of oxytocin in blood draining the ovary exceeded the arterial concentration in samples taken before administering Estrumate by a factor of 2.8 ($n = 12$; $P < 0.001$), indicating basal ovarian secretion of oxytocin.

These observations show that oxytocin is released into the ovarian vein both before and after prostaglandin treatment; thus oxytocin is secreted by the ovary. It is not certain, however, whether the oxytocin released is synthesized in the ovary or sequestered in it from the blood stream (for example, bound to a receptor on the plasma membrane). The demonstration that oxytocin is released from the ovary before Estrumate administration must be interpreted with caution, as manipulation of the uterus during insertion of catheters may have caused sufficient uterine secretion of prostaglandin $F_{2\alpha}$ to stimulate ovarian release of oxytocin[9].

Assuming a minimum ovarian venous blood flow of 10 ml min^{-1}, and correcting ovarian venous concentrations for those in arterial blood, it can be calculated that during the 40 min following prostaglandin treatment average ovarian release of oxytocin was 0.24 μg per corpus luteum. This is approximately 20% of the luteal content of oxytocin reported by Wathes and Swann[8], and therefore it appears that a large proportion of the oxytocin contained in the ovary is available for release in response to prostaglandin stimulation. Although in the present experiments we used an analogue of prostaglandin $F_{2\alpha}$ to stimulate ovarian oxytocin secretion, it seems likely that oxytocin is also secreted in response to endogenous prostaglandin $F_{2\alpha}$ released from the uterus during the oestrous cycle. Synchronous surges of oxytocin-neurophysin, or of oxytocin itself, and of 13,14-dihydro-15-ketoprostaglandin $F_{2\alpha}$ (pulmonary metabolite of prostaglandin $F_{2\alpha}$) have been observed in sheep (ref. 10 and A.P.F.F. and E.L.S., unpublished) and oxytocin secretion has been reported in other species in response to native prostaglandin $F_{2\alpha}$[11,12].

Ovarian oxytocin may be involved in limiting luteal progesterone secretion by a local mechanism: oxytocin has been shown to inhibit steroidogenesis in ovarian[13] and testicular[14] tissues. It may also act systemically to augment the uterine secretion of prostaglandin $F_{2\alpha}$[15], thereby ensuring rapid completion of luteal involution, or lead to the reduction in luteal blood flow which occurs at luteolysis[16]: oxytocin has been reported to show vasoconstrictor properties in the mammary gland in some conditions[17]. However, as in so many cases of well established hormones being discovered in novel places[18], the physiological role of ovarian oxytocin remains to be determined.

Received 16 March; accepted 13 April 1982.

1. Armstrong, D. T. & Hansel, W. *J. Dairy Sci.* **42**, 533–542 (1959).
2. Milne, J. A. *Aust. vet. J.* **39**, 51–52 (1963).
3. Cooke, R. G. & Knifton, A. *Theriogenology* **16**, 95–97 (1981).
4. Sheldrick, E. L., Mitchell, M. D. & Flint, A. P. F. *J. Reprod. Fert.* **59**, 37–42 (1980).
5. Webb, R., Mitchell, M. D., Falconer, J. & Robinson, J. S. *Prostaglandins* **22**, 443–453 (1981).
6. Sheldrick, E. L. & Flint, A. P. F. *Prostaglandins* **22**, 631–636 (1981).
7. Schams, D., Lahlou-Kassi, A. & Glatzel, P. *J. Endocr.* **92**, 9–13 (1982).
8. Wathes, D. C. & Swann, R. W. *Nature* **297**, 225–227 (1982).
9. Roberts, J. S., Barcikowski, B., Wilson, L., Skarnes, R. C. & McCracken, J. A. *J. Steroid Biochem.* **6**, 1091–1097 (1975).
10. Fairclough, R. J. *et al. Prostaglandins* **20**, 199–208 (1980).
11. Ellendorff, F. *et al. Br. J. Pharmac.* **62**, 412P (1978).
12. Gillespie, A., Brummer, H. C. & Chard, T. *Br. med. J.* **1**, 543–544 (1972).
13. Tan, G. J. S., Tweedale, R. & Biggs, J. S. G. *J. Steroid Biochem.* **14**, xiii (1981).
14. Adashi, E. Y. & Hsueh, A. J. W. *Nature* **293**, 650–652 (1981).
15. McCracken, J. A. *Adv. Prostaglandin Thromboxane Res.* **8**, 1329–1344 (1980).
16. Pharriss, B. B., Cornette, J. C. & Gutknecht, G. D. *J. Reprod. Fert. Suppl.* **10**, 97–103 (1970).
17. Linzell, J. L. in *Lactation* Vol. 1 (eds Larson, B. L., & Smith, V. R.) 143–225 (Academic, New York, 1974).
18. Sharpe, R. M. *J. Reprod. Fert.* **64**, 517–527 (1982).

Racial differences in hypertension-associated red cell sodium permeability

Nina L. Etkin*, John R. Mahoney†, Michael W. Forsthoefel‡, James R. Eckman‡, John D. McSwigan†, Richard F. Gillum§ & John W. Eaton*†§

Departments of *Anthropology, †Laboratory Medicine and Pathology, and §Medicine, University of Minnesota, Minneapolis, Minnesota 55455, USA
‡ Department of Medicine, Emory University, Atlanta, Georgia 30322, USA

The prevalence and severity of essential hypertension varies greatly among human populations. Some of this variation is undoubtedly due to differences in dietary Na^+ levels[1]. However, even within groups which consume large amounts of salt, not all members develop increased blood pressure. Furthermore, the clinical expression of hypertension varies among races, and American blacks have particularly severe symptoms. These observations, in addition to a familial pattern in the occurrence of essential hypertension, suggest important genetic influences. Recent studies, primarily of northern European populations, have demonstrated abnormal erythrocyte Na^+ transport in affected individuals[2–5] and in approximately half the members of kindreds founded by a hypertensive progenitor[4,5]. We have found that red cell Na^+ transport can be assessed simply by measuring the unidirectional passive influx of $^{22}Na^+$ into ouabain-treated erythrocytes[5] and have applied this technique to American blacks with essential hypertension. The results suggest a complete absence of the abnormal erythrocyte Na^+ transport which is characteristic of white patients with essential hypertension. Thus, among American blacks, essential hypertension may have either a different genetic basis or a substantially different expression.

The white and black patients investigated had had supine diastolic blood pressures exceeding 95 mm Hg for at least 2 yr, and secondary hypertension was excluded by normal renal function and (in most cases) normal intravenous pyelogram. Normotensive controls had supine diastolic pressures of <85 mm Hg and negative family histories for essential hypertension. All patients were haemoglobin AA, as determined by isoelectric focusing.

Figure 1 shows the values obtained when erythrocyte Na^+ influx was measured in the four groups studied. In 21 patients with essential hypertension, erythrocyte $^{22}Na^+$ influx averaged twice that into red cells from 21 normotensive white controls (0.527 ± 0.128 compared with 0.266 ± 0.044 mmol Na^+ per l red cells per h respectively; $t = 8.87$, $P < 0.001$). The values for only 2 of 21 white patients with essential hypertension overlap the upper range of normal. This increased Na^+ influx is not influenced by antihypertensive medications administered to the patient[5] and is not caused by elevated blood pressure *per se*; erythrocytes from 12 patients with secondary hypertension (caused by renal disease) had values within the normal range (0.233 ± 0.060 mmol Na^+ per l red cells per h).

In contrast, red cells from 32 American black patients with essential hypertension showed no evidence of enhanced Na^+ influx. In these patients, values for erythrocyte Na^+ influx (0.202 ± 0.059 mmol Na^+ per l red cells per h) were indistinguishable from those for 23 normotensive black controls (0.216 ± 0.069 mmol Na^+ per l red cells per h; $t = 0.79$, $P > 0.10$). There is, however, a highly significant difference between hypertensive whites and blacks in erythrocyte Na^+ influx ($t = 10.66$, $P < 0.001$). Limited *in vitro* studies indicate that loop diuretics such as furosemide (10^{-3} M) preferentially decrease net Na^+ influx in hypertensive whites while having little

0028-0836/82/240588—02$01.00

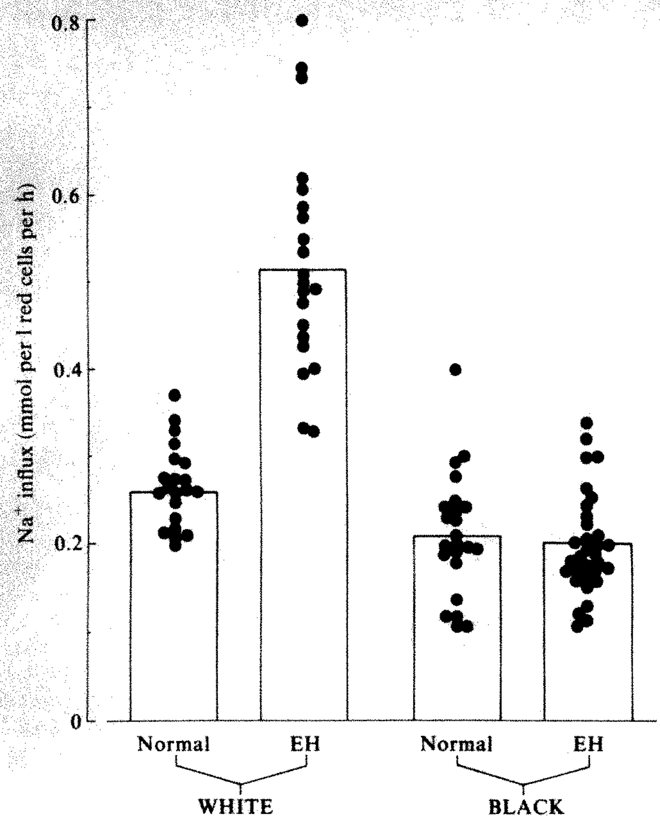


Fig. 1 Comparison of red cell Na^+ influx in normal and essential hypertensive (EH) whites and blacks. Each point represents the value for a single individual and the bars designate the mean for each group. Samples of venous blood were collected in heparinized vacutainers. Some of these were shipped on ice, with a maximum storage time of 21 h. In these cases, we included both shipment- and time-control samples from normal and hypertensive patients. The preparation of erythrocytes and assessment of $^{22}Na^+$ influx were performed as previously reported[5], except that stored samples of whole blood were re-equilibrated for 1 h at 37 °C. Red cells were washed in a buffer comprising 145 mM KCl, 5 mM NaCl, 10 mM Tris, 10 mM morpholinopropanesulphonic acid (MOPS), 10 mM D-glucose and 0.1 mM ouabain, *p*H 7.40 at 37 °C. 0.5 ml of washed, packed erythrocytes of known haematocrit (~75 vol %) was admixed with 0.5 ml of the same buffer containing 0.1 μCi of $^{22}Na^+$ (final specific activity ~20 μCi $mmol^{-1}$; NEN). The cell suspension was incubated in a shaking water bath at 37 °C for 30 min, then the cells were rapidly washed three times in 5 vols of ice-cold buffer containing 75 mM $MgCl_2$, 85 mM sucrose, 10 mM D-glucose, 10 mM Tris and 10 mM MOPS, *p*H 7.40. 0.3 ml of washed, packed cells was precipitated with 0.6 ml of 10% trichloroacetic acid. Supernatant radioactivity was assessed by liquid scintillation counting, and Na^+ influx was expressed as mmol Na^+ per l red cells per h. All determinations were made in duplicate, and the resultant values for each individual agree to within ±5%.

quantitative effect on the other three groups.

These observations suggest that either hypertension has a different genetic aetiology in the two groups or its expression (as reflected by variations in passive Na^+ transport) differs greatly. Indeed, hypertension among American blacks is also characterized by, a higher prevalence, younger age of onset and elevated risk of mortality; relatively normal renin and dopamine β-hydroxylase activities; more difficult therapeutic control; and an increased severity of secondary manifestations such as nephrosclerosis, retinopathy, accelerated atherogenesis and other vascular diseases[6-8]. The existence of a significant disparity between hypertensive whites and blacks in red cell Na^+ transport strengthens the clinical impression that hypertension in black patients differs qualitatively from that in whites. These variations in red cell Na^+ transport may be important in helping unravel the genetic basis of essential hypertension and its variable incidence and expression in different populations.

This work was supported in part by grants from the American Heart Association (81-715), the Wyoming and Montana AHA affiliates and NIH (HL-16833). N.L.E. was supported by a Faculty Research Appointment from the University of Minnesota, J.R.M. is recipient of a G. T. Evans Fellowship and J.W.E. is supported by an NIH Research Career Development Award.

Received 31 December 1981; accepted 16 April 1982.

1. Freis, E. D. *Circulation* **53**, 589–595 (1976).
2. Canessa, M., Adragna, N., Solomon, H. S., Connolly, T. M. & Tosteson, D. C. *New Engl. J. Med.* **302**, 772–776 (1980).
3. Garay, R. P., Elghozi, J.-L., Dagher, G. & Meyer, P. *New Engl. J. Med.* **302**, 769–771 (1980).
4. Meyer, P. *et al. Br. Med. J.* **282**, 1114–1117 (1981).
5. Mahoney, J. R., Etkin, N. L., McSwigan, J. D. & Eaton, J. W. *Blood* **59**, 439–442 (1982).
6. Saunders, E. & Williams, R. A. in *Textbook of Black-Related Diseases* (ed. Williams, R. A.) 333–358 (McGraw-Hill, New York, 1975).
7. Voors, A. W., Berenson, G. S., Dalferes, E. R., Webber, L. S. & Shuler, S. E. *Science* **204**, 1091–1094 (1979).
8. Warren, S. E. & O'Connor, D. R. *Am. J. Med.* **69**, 425–429 (1980).

B-cell subpopulations identified by two-colour fluorescence analysis

R. R. Hardy, K. Hayakawa, J. Haaijman & L. A. Herzenberg

Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA

The simultaneous and rapid measurement of the amounts of two different fluorochrome-coupled antibodies bound to single cells (two-colour immunofluorescence) provides a very powerful means for the identification of lymphocyte subpopulations[1]. Using a dual-laser fluorescence-activated cell sorter (FACS) we show that two monoclonal antibodies, anti-IgM and anti-IgD, labelled respectively with fluorescein and 'Texas red' (a new red-fluorescent dye) reveal several previously unrecognized B-cell subpopulations in mouse spleen and lymph nodes. Measured individually, these surface markers (IgM and IgD) show only that B cells are broadly heterogeneous with respect to the amount of surface immunoglobulin expressed[2]; however, measured simultaneously, they clearly define at least two B-cell subsets. One of these populations, which is predominant in spleen and constitutes the overwhelming majority of B cells in lymph nodes, is missing in CBA/N (Xid) mice known to be deficient with respect to their B-cell immune responses[3].

The B-cell populations in BALB/c spleen (Figs 1, 2) are resolved into two subsets on the basis of the amount of IgM expressed. The predominant population (labelled I in the diagram), which has relatively little surface IgM, expresses intermediate to high levels of surface IgD and constitutes ~60% of the splenic B cells (Table 1). This population is even more predominant in BALB/c lymph nodes where it constitutes >80% of the B cells.

The remaining BALB/c splenic population(s), which express 10–100 times more IgM, appear to be further subdivided according to the amount of surface IgD expressed (labelled II and III) and the organ localization of the cells. The 'high-IgD, high-IgM' population (II) constitutes ~20% of the B cells in spleen and lymph nodes. The 'low-IgD' population (III) is distinguishable more by its apparent absence from lymph nodes (see Fig. 1) than by any clear-cut demarcation in the staining pattern.

Low-angle light scatter measurements in the FACS, which provide an index of cell size[4], further distinguish these populations. Population I is the major stained component of the smaller (by scatter) cells; population II is found predominantly among the larger cells; and population III is found amongst cells in an intermediate size range. This scatter difference is another parameter which should help in isolating these subpopulations for use in functional assays.

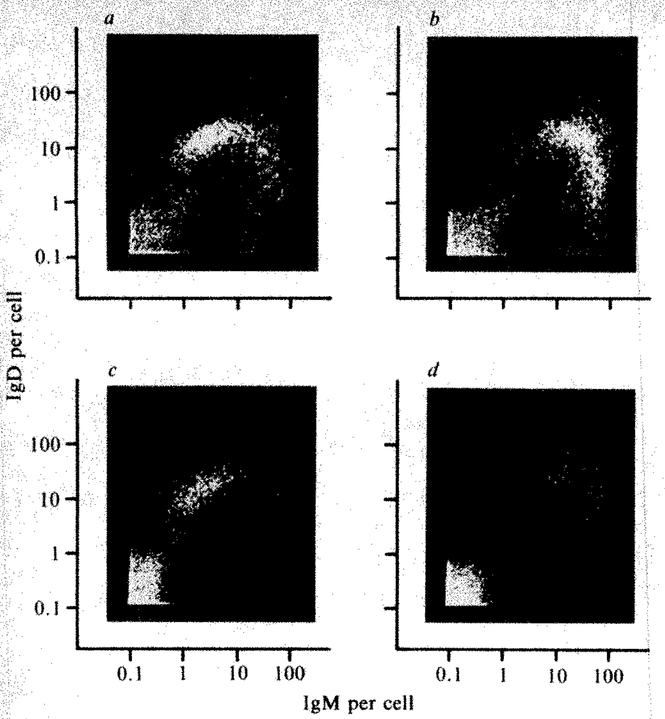


Fig. 1 B-cell subpopulations defined by two-colour FACS analysis of cells stained simultaneously with anti-IgM and anti-IgD antibodies. The individual points in the 'dot plots'[14] shown represent the amounts of IgD and IgM expressed on 3,000 individual cells from spleen and lymph nodes in CBA/N and BALB/c mice. The relative amounts of IgM and IgD per cell are indicated on the axes. *a*, BALB/c spleen; *b*, CBA/N spleen; *c*, BALB/c peripheral lymph nodes; *d*, CBA/N peripheral lymph nodes. Spleen or lymph node cells (10^6) were stained with 0.5 μg of fluorescein-labelled monoclonal rat anti-IgM[15] and 0.5 μg of biotinated monoclonal mouse anti-IgD[16] antibodies in 100 μl of biotin-free RPMI-1640 containing 10 mM HEPES buffer, 0.1% sodium azide and 3% newborn calf serum for 30 min at 0 °C. Cells were washed three times with RPMI and stained with 1 μg of Texas red–avidin in 50 μl of RPMI for 30 min at 0 °C. Cells were washed three times with RPMI, resuspended in 300 μl of the same buffer and analysed on the dual-laser FACS equipped with logarithmic amplifiers. To permit subsequent analyses, 'list-mode' data recording the scatter and two fluorescence measurements for each cell were collected on 30,000 cells using a VAX 11/780 computer.

Studies with CBA/N mice further clarify the definition of the B-cell subpopulations. These mice, which carry an X-linked B-cell defect, have almost no detectable population I cells in lymph nodes or spleen. Their splenic B cells, which consist entirely of high-IgM populations, are divided between cells comparable with the cells in populations II and III in BALB/c spleen. The CBA/N spleen also has many more IgM-, IgD-double negative cells than the BALB/c spleen. As might be expected, CBA/N spleen cells have the same scatter correlations as BALB/c spleen cells, except that the small cells lack both IgM and IgD.

There are only a few immunoglobulin-positive cells in CBA/N lymph nodes and these stain exclusively in the region of population II. Thus, when population I is absent from lymph node and consequently does not obscure the analysis of the minor B-cell populations present, populations II and III are distinguishable by their locations in the IgM, IgD fluorescence plots.

Thus the combined data from two-colour FACS analyses of CBA/N and BALB/c spleen and lymph node demonstrate that the heterogeneous expression of IgM and IgD on the whole B-cell population reflects three (or more) subpopulations, each with characteristic levels of surface immunoglobulins. Studies of the appearance of these populations during development, and the disappearance of the B-cell population(s) present predominantly during neonatal life, indicate the presence of at least one more B-cell subpopulation characterized by the lack of surface IgD (J.H., R.R.H., K.H. and L.A.H. in preparation). While it is clear that population I reaches mature level last[5] the relationship between these populations is currently unknown. An allotype-specific IgM–IgD stain carried out on sorted cells transferred to allotype congenic mice should help to clarify these relationships.

Table 1 Per cent of cells in IgM/IgD defined B-cell subpopulations

Strain	Organ	I	II	III
BALB/c	Spleen	30	10	10
	Lymph node	17	4	<1
CBA/N	Spleen	<4*	20	15
	Lymph node	<1	4	<1
CBA	Spleen	23	18	8
	Lymph node	10	4	<1

CBA/N mice lack the major B-cell subpopulation found in spleen and lymph nodes in normal mice. Subpopulations were defined by two-colour FACS analysis (Fig. 2). Data show the per cent of total spleen or lymph node cells present in the B-cell subpopulations shown in Fig. 2. Percentages for populations I and II may be in error by as much as 5% in the BALB/c and CBA determinations due to overlap between these populations.

*This value (4%) is attributable to overlap from population II.

The absence of population I in CBA/N mice fits with several previous observations: the higher IgM/IgD ratio on B cells from CBA/N spleen[6] arises from a lack of the low IgM/IgD ratio (population I) cells in these mice; the absence of certain B cell determinants [Lyb3 (ref. 7), Lyb5 (ref. 8), Lyb7 (ref. 9)] in CBA/N is apparently due to the presence of these markers exclusively on population I cells; the presence of one of these markers (Lyb3) on virtually all normal lymph node B cells, but on only a fraction of normal spleenic B cells[10], is consistent with the relative amounts of population I in the two organs.

CBA/N mice have an extensively studied immune deficiency: they do not make antibody to type II thymus-independent antigens (such as dinitrophenyl–Ficoll)[11]; their spleen cells fail to give the usual *in vitro* proliferative responses shown by normal spleen cells stimulated with anti-immunoglobulin antisera[12]; and their serum IgM and IgG3 levels are very low compared with normal mice[13]. We speculate that the B cells responding to dinitrophenyl–Ficoll or to anti-immunoglobulin antisera in the proliferative assay are contained in subpopulation I (missing in CBA/N). The B cells responsible for a large part of IgM and IgG3 antibody production also may be included in this subset. We should be able to test these hypotheses and assign functions to the B-cell populations defined here by

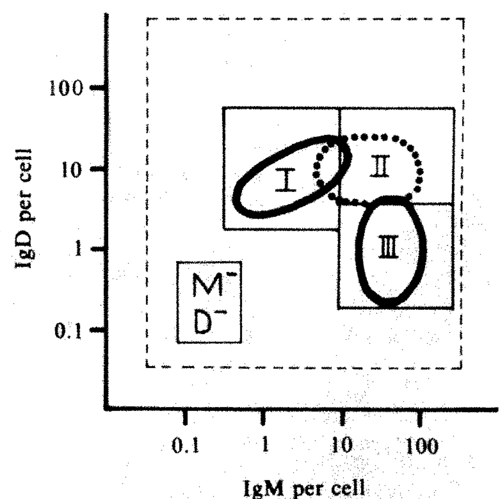


Fig. 2 Diagram of the B-cell subpopulations shown in Fig. 1 (see text). The integration bounds used to obtain the percentages in Table 1 are also shown.

sorting these populations before appropriate *in vitro* and adoptive transfer assays.

We acknowledge the Institute of Experimental Gerontology, Health Organization TNO, The Netherlands, for providing CBA/N mice. We thank Dr P. Kincade for the monoclonal rat anti-IgM antibody, Dr D. Parks for invaluable assistance with the dual-laser FACS and Mr W. Moore for writing the VAX data handling programs. This work was supported in part by NIH grants CA-04681 and GM-17367.

Received 31 December 1981; accepted 15 April 1982.

1. Ledbetter, J. A., Rouse, R. V., Micklem, H. S. & Herzenberg, L. A. *J. exp. Med.* **152**, 280–295 (1980).
2. Vitetta, E. S. *et al.* in *Progress in Immunology* Vol. 3 (ed. Mandel, T.) 65–73 (North-Holland, Amsterdam, 1977).
3. Scher, I., Frantz, M. M. & Steinberg, A. D. *J. Immun.* **110**, 1396–1401 (1973).
4. Fathman, C. G., Small, M., Herzenberg, L. A. & Weissman, I. L. *Cell. Immun.* **15**, 109–128 (1975).
5. Hardy, R. R., Hayakawa, K. & Herzenberg, L. A. *Ann. N.Y. Acad. Sci.* (in the press).
6. Finkelman, F. D., Smith, A. H., Scher, I. & Paul, W. E. *J. exp. Med.* **142**, 1316–1321 (1975).
7. Huber, B., Gershon, R. K. & Cantor, H. *J. exp. Med.* **145**, 10–20 (1977).
8. Ahmed, A. *et al. J. exp. Med.* **145**, 101–110 (1977).
9. Subbarao, B. *et al. J. Immun.* **122**, 2279–2285 (1979).
10. Huber, B., Gathings, W. E. & Cooper, M. D. in *B Lymphocytes in the Immune Response* (eds Cooper, M. D., Mosier, D. E., Scher, I. & Vitetta, E.) 125–130 (Elsevier, Amsterdam, 1979).
11. Mosier, D. E., Mond, J. J. & Goldings, E. A. *J. Immun.* **119**, 1874–1878 (1977).
12. McKenzie, I. F. C. & Potter, T. *Adv. Immun.* **27**, 241 (1979).
13. Perlmutter, R. M. *et al. J. exp Med.* **149**, 993–998 (1979).
14. Herzenberg, L. A. & Herzenberg, L. A. in *Handbook of Experimental Immunology* 3rd edn (ed. Weir, D. M.) 22.1–22.21 (Blackwell Scientific, Oxford, 1978).
15. Kincade, P. W., Lee, G., Sun, L. & Watanabe, T. *J. Immun. Meth.* **42**, 17–26 (1981).
16. Oi, V. T., Jones, P. P., Goding, J. W., Herzenberg, L. A. & Herzenberg, L. A. *Curr. Topics Microbiol. Immun.* **8**, 115–129 (1978).

Target antigens of purified human immunoglobulins which inhibit growth of *Plasmodium falciparum in vitro*

Graham V. Brown*, Robin F. Anders*, Graham F. Mitchell* & Peter F. Heywood†

* Laboratory of Immunoparasitology, The Walter and Eliza Hall Institute of Medical Research, Melbourne, Victoria 3050, Australia
† The Papua New Guinea Institute of Medical Research, Madang, Papua New Guinea

A prerequisite for an antigenically defined vaccine against malaria is the identification of parasite antigens which are necessary or sufficient for the induction and expression of host-protective immunity. As malaria infection in man can be treated by passive transfer of serum or immunoglobulin from adults in an endemic area[1] it is reasonable to propose that sera from clinically defined patient groups and functionally defined immunoglobulin may be used as probes to identify the relevant parasite antigens involved in host protection ('host-protective antigens')[2]. We have now purified immunoglobulin from a large number of individuals living in the endemic area, tested for inhibitory effects on *in vitro* growth of *Plasmodium falciparum* originating from the same area, then examined antibody specificities by immunoprecipitation analysis with biosynthetically labelled proteins of blood stages of the parasite. Two-dimensional gel analysis of immunoprecipitates revealed a correlation between inhibition of parasite growth and increased antibody binding to two proteins of molecular weight (M_r) ~96,000.

Serum was obtained with informed consent from healthy adult blood donors living in the Madang region of Papua New Guinea (PNG), where malaria is endemic, at a time of the year when transmission was low. Purified immunoglobulin from each donor (prepared using protein A–Sepharose, Pharmacia) was screened for its ability to inhibit the *in vitro* growth of a PNG isolate of *P. falciparum* by a modification of a method described previously[3]. In all experiments we used a single parasite isolate which has been maintained in our laboratory since it was obtained from an infected child in Madang in 1979. Briefly, immunoglobulin was dialysed against culture medium (RPMI 1640 with sodium bicarbonate and HEPES buffer) for 36 h and adjusted to a final concentration of 5 mg ml^{-1} in culture medium. After addition of normal human serum to a final concentration of 10%, the 'immunoglobulin-supplemented medium' was assayed for its ability to inhibit the growth of *P. falciparum in vitro*. In this assay, 200 μl of parasitized blood group O erythrocytes in culture medium (8% v/v) were added to a series of flat-bottomed wells of a microtitre tray and cultured in a candle jar using the method devised by Trager and Jensen[4]. Each day, culture medium was changed (using normal or test medium) until day 4, when parasitaemia was assessed by uptake of ^{3}H-leucine into trichloracetic acid-precipitable material, such uptake having been shown to correlate with the percentage of parasitized cells (G.V.B., unpublished). Using this assay, immunoglobulin prepared from 4 of 39 sera inhibited growth by more than 50%.

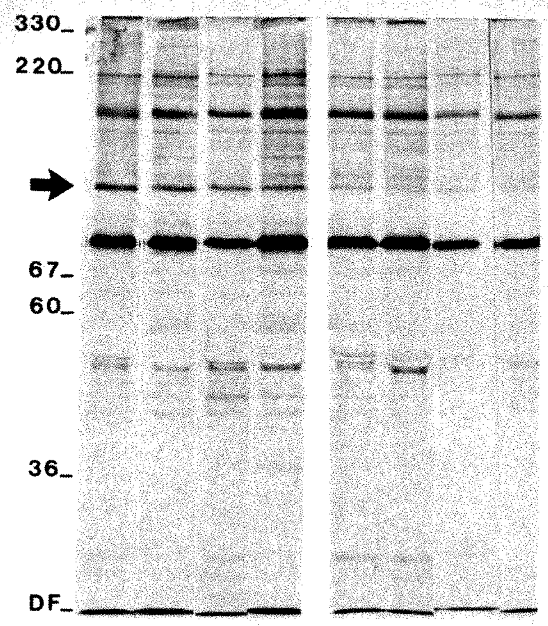


Fig. 1 SDS-polyacrylamide gel electrophoresis of immunoprecipitates of detergent extracts of ^{35}S-methionine-labelled *P. falciparum* antigen with protein A-purified immunoglobulin from adults living in the endemic area. The arrow (→) indicates a band which is dominant in immunoprecipitates with the four most inhibitory immunoglobulins (left-hand lanes), compared with four non-inhibitory immunoglobulins (right-hand lanes). Molecular weight markers are given (×10^{-3}). DF, dye front.

Pools were made of five sera containing immunoglobulin which inhibited parasite growth and five sera containing immunoglobulin which, at the same concentration (5 mg ml^{-1}), had no inhibitory effect on parasite growth. IgG was prepared from them using a different method from that used for the individual samples, by precipitation with 45% ammonium sulphate followed by ion exchange chromatography on DEAE-cellulose. The IgG was then dialysed against culture medium and assayed for inhibition of *P. falciparum* growth *in vitro*. The pooled inhibitory IgG caused at least 65% inhibition of parasite growth compared with non-inhibitory IgG (50,800 ± 420 c.p.m. for non-inhibitory IgG compared with 17,150 ± 750 c.p.m. for inhibitory IgG using the radioisotopic readout for parasite growth inhibition).

Biosynthetically labelled parasite antigens were prepared by culturing the same isolate of PNG *P. falciparum* as used for the inhibition studies for 14 h in methionine-free medium (RPMI 1640 Select-Amine Kit, Gibco) supplemented with ~200 μCi ml^{-1} ^{35}S-methionine (800 Ci mmol^{-1}, Radiochemical Centre). A detergent extract of parasitized cells was made using 0.5% Triton X-100 in NaCl, EDTA and Tris buffer[5] (Triton NET) containing methionine (2 mg ml^{-1}). After centrifugation at 25,000*g* for 30 min, the supernatant was passed over a Sephadex G-25M column (PD10 column, Pharmacia) and eluted in 0.5% Triton NET supplemented with methionine.

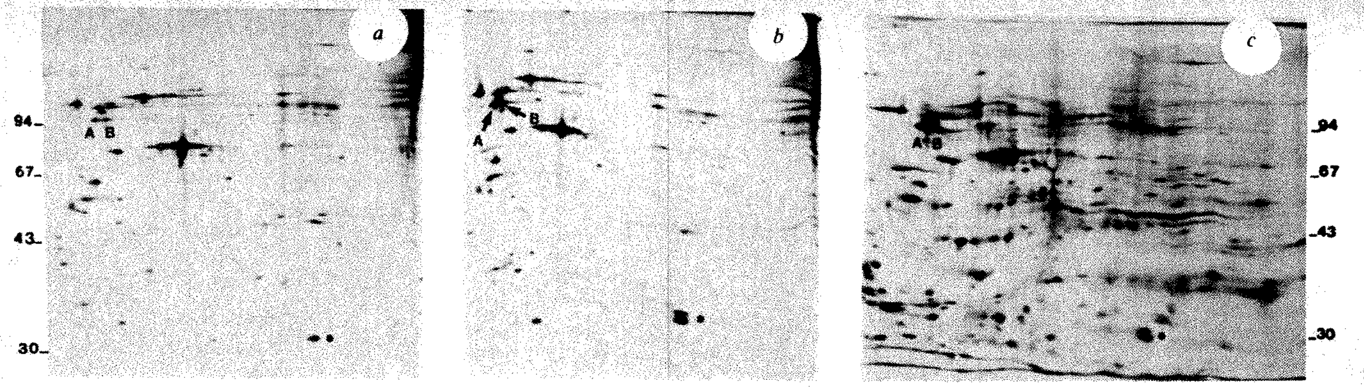


Fig. 2 Immunoprecipitates of detergent extracts of ^{35}S-methionine-labelled *P. falciparum* (*c*) by purified human immunoglobulin which had no effect on parasite growth (*a*) or inhibited parasite growth (*b*). Inhibitory IgG contained a high titre of antibody to the two acidic proteins, M_r 96,000 labelled A, B. Before isoelectic focusing, the antigen–antibody complexes bound to *S. aureus* protein A were solubilized in 2% w/v SDS lysis buffer containing 9.5 M urea, 2% ampholines (4:1 v/v *p*H 5–8:*p*H 3.5–10 ampholines, LKB) and 50 mM dithiothreitol. After 20 min at room temperature, five times the volume of lysis buffer containing 2% Triton X-100, 2% ampholines, 9.5 M urea and 50 mM dithiothreitol was added. Incubation was continued for 15 min at room temperature before samples were centrifuged before analysis, or if necessary, stored at −70 °C. For the second dimension, electrophoresis in reducing conditions in the presence of SDS was performed on 10% acrylamide slab gels using a discontinuous buffer system[6]. Gels were prepared for fluorography using EN^{3}HANCE (NEN) before drying down and exposure to Kodak X-Omat S film. Numbers refer to the molecular weights ($\times 10^{-3}$) of M_r standards (Pharmacia Fine Chemicals Low Molecular Weight Calibration Kit) which were run on each slab and identified by Coomassie blue staining. Acidic region is to the left of each gel.

IgG isolated from each pool of sera, or protein A-purified immunoglobulin from individual sera, were reacted with biosynthetically labelled parasite proteins and antigen–antibody complexes precipitated with *Staphylococcus aureus* as described previously[3] before analysis by one[6] and two-dimensional gel electrophoresis[7]. Approximately equal amounts of radioactivity were precipitated from preparations of solubilized biosynthetically labelled infected cells by IgG from the two pools of sera. Examination of the fluorographs (Fig. 1) showed that the majority of proteins precipitated by non-inhibitory immunoglobulins (right-hand lanes) were precipitated to the same extent by the inhibitory immunoglobulins (left-hand lanes). However, one band of M_r 96,000 (indicated by the arrow) was dominant in immunoprecipitates using inhibitory immunoglobulins but was poorly precipitated by non-inhibitory immunoglobulins. Two-dimensional analysis of immunoprecipitates using IgG from sera pools confirmed that there were many similarities in proteins precipitated by non-inhibitory IgG (Fig. 2*a*) and inhibitory IgG (Fig. 2*b*). However, as in the one-dimensional analysis there was a difference in the two-dimensional gel patterns at M_r 96,000 in that two acidic proteins (labelled A, B in Fig. 2) were much more prominent in precipitates with inhibitory IgG. Different antigen preparations made from the same isolate of *P. falciparum* contained varying proportions of different parasite proteins[8], presumably because the proportion of parasites at each stage in the life cycle in asynchronous culture varied each time an antigen preparation was made. For example, other antigen preparations of the same isolate (not shown) had a greater proportion of high molecular weight proteins. Antigen A (Fig. 2) was invariably well represented while antigen B was sometimes present in only small amounts. The representation of several minor proteins also differed between the immunoprecipitates, including a protein with approximately the same isoelectric point (*p*I) as the M_r 96,000 proteins but of M_r ~40,000. Another difference in precipitation of a low molecular weight protein (* in Fig. 2) has not been emphasized because it was not reproducible.

We thus conclude that natural infection in man leads to antibody production which can inhibit the growth of *P. falciparum in vitro*. We found that only 4 of 39 immunoglobulin preparations inhibited parasite growth by >50% while some samples (9 of 39) actually increased growth beyond control levels (a result consistent with previous studies using unfractionated sera[9]). Thus, it seems that the effects of immunoglobulin on parasite growth *in vitro* may be a balance between growth promoting and inhibiting activities. In the present study, the population donating sera at a time of the year when transmission was low contained relatively few individuals whose serum had net inhibitory activity.

IgG which inhibited parasite growth immunoprecipitated two proteins (M_r 96,000) more efficiently than did IgG with no inhibitory effect on parasite growth. In previous studies using one-dimensional gel analysis to compare immunoprecipitates of inhibitory and non-inhibitory serum[3], we noted differences in recognition by the sera of a protein band of approximately the same molecular weight. Thus we have identified at least two proteins (designated A and B), antibody to which correlates with *in vitro* inhibition of growth of *P. falciparum*. The observations strongly suggest that these protein antigens may be important in the induction (and expression) of host-protective immunity in man as a result of exposure to infection in an endemic area for malaria. We have examined the antigenic composition of 15 other isolätes of *P. falciparum* from Papua New Guinea and all contain antigen A as a major biosynthetically labelled protein.

Other investigators have examined the effect of human[9] or immune monkey[10–12] sera on the growth of *P. falciparum in vitro* but have been unable to correlate the inhibition with antibody to specific antigens. Hybridoma-derived murine monoclonal antibodies which inhibit the growth of *P. falciparum in vitro* have been reported recently[13]. One of these hybridoma antibodies reacted with two polypeptides of M_r 96,000 and 36,000 (ref. 13). Our finding that purified human IgG which inhibits parasite growth *in vitro* has excess antibody directed against two proteins of M_r 96,000 suggests that the hybridoma-derived antibody has the same specificity as that of antibodies contained in human IgG preparations which block parasite growth.

Additional proteins must be present in the blood stages of *P. falciparum* capable of inducing a host protective immune response in man which have not been demonstrated by the techniques we have outlined. Some immunoglobulin preparations causing 30–50% inhibition of parasite growth *in vitro* did not precipitate relatively large amounts of antigens A and B, confirming that other specificities are also important. Our techniques would not detect proteins with little methionine, proteins not solubilized by detergent extraction or after immune complex precipitation, very basic proteins, or non-protein antigens. Note also that other correlates with protection may be important at times of high transmission with multiple antigenic variants, high parasite loads and circulating immune complexes, and with the

possibility of rise and fall of competing facilitatory and inhibitory antibodies. Furthermore, analysis of other mechanisms in which immunoglobulin may have a host-protective role, such as opsonization, may reveal yet other antigenic correlates. It remains to be determined whether the proteins identified by this comparative immunoprecipitation approach are immunogenic after isolation, and whether they induce host protection or inhibitory antibodies in appropriate model systems.

The work was supported by the National Health and Medical Research Council of Australia, the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases and the Rockefeller Foundation Great Neglected Diseases of Mankind Network. We acknowledge the technical assistance of Raelene Grumont.

Received 23 November 1981; accepted 14 April 1982.

1. Cohen, S., McGregor, I. A. & Carrington, S. C. *Nature* **192**, 733–737 (1961).
2. Knopf, P. M., Brown, G. V., Howard, R. J. & Mitchell, G. F. *Aust. J. exp. Biol. med. Sci.* **57**, 603–615 (1979).
3. Brown, G. V., Anders, R. F., Stace, J. D., Alpers, M. P. & Mitchell, G. F. *Parasite Immun.* **3**, 283–298 (1982).
4. Jensen, J. B. & Trager, W. *J. Parasit.* **63**, 883–886 (1977).
5. Kessler, S. W. *J. Immun.* **115**, 1482–1490 (1975).
6. Laemmli, U. K. & Favre, M. *J. molec. Biol.* **80**, 575–599 (1973).
7. O'Farrell, P. H. *J. biol. Chem.* **250**, 4007–4021 (1975).
8. Brown, G. V., Coppel, R. L., Vrbova, H., Grumont, R. J. & Anders, R. F. *Exp. Parasit.* **53**, 279–284 (1982).
9. Wilson, R. J. M. & Phillips, R. S. *Nature* **263**, 132–134 (1976).
10. Reese, R. T. & Motyl, M. R. *J. Immun.* **123**, 1894–1899 (1979).
11. Butcher, G. A. & Cohen, S. *Immunology* **23**, 503–519 (1972).
12. Campbell, G. H., Mrema, J. E. K., O'Leary, T. R., Jost, R. C. & Rieckman, K. H. *Bull. Wld Hlth Org.* **57**, Suppl. 1, 219–215 (1979).
13. Perrin, L. H., Ramirez, E., Lambert, P. H. & Miescher, P. A. *Nature* **289**, 301–303 (1981).

Cancer metastasis is selective or random depending on the parent tumour population

J. E. Talmadge & I. J. Fidler

Cancer Metastasis and Treatment Laboratory, NCI-Frederick Cancer Research Facility, PO Box B, Frederick, Maryland 21701, USA

The issue of whether metastases result from the random survival of cells released from the primary tumour or from the selective growth of specialized subpopulations of cells having properties that allow them to complete the metastatic process is important to our understanding of tumour biology[1,2]. Previous studies have provided indirect evidence that the process of metastasis favours the survival of cells having a metastatic phenotype(s)[3–8]. Direct evidence that the process is selective would be provided by the demonstration that cells populating spontaneous metastases are, in general, more metastatic than most of the parent tumour cells; some[9–11], but not all[12,13], previous reports have failed to provide such evidence, suggesting that the process is random. These discrepancies could be due to differences in the biological characteristics of the various tumour systems studied and in the experimental conditions[14]. In the present study, these variables were minimized by using three metastatic variant cell lines of the B16 melanoma. We report that, even in controlled conditions, the process of metastasis can appear as either selective or random depending on the nature of the initial population. Specifically, metastasis by the unselected, poorly metastatic parent B16 tumour was indeed selective. In contrast, metastasis by previously selected B16 lines appeared to be random.

To determine whether metastases contain cells having increased metastatic potential, we used three metastatic tumour cell variants isolated from the B16 melanoma. B16-F1 is an unselected tumour cell line that metastasizes poorly when tumour cells are implanted intravenously (i.v.; experimental metastasis) or subcutaneously (s.c.; spontaneous metastasis)[15]. The B16-F10 cell line[15] was selected for its ability to colonize the lung after i.v. administration, and the B16-BL6 cell line[16] was selected *in vitro* for its invasiveness, demonstrated by a high incidence of spontaneous metastases after s.c. implantation. All three tumour cell lines were implanted into the foot-pads of syngeneic 6-week-old C57BL/6 female mice (20 per group). After 5–7 weeks, when one or two mice from each group became listless, the entire group was killed. Only 5 of 20 mice injected with B16-F1 cells exhibited lung metastases (median number of lung colonies 0, range 0–12); 8 of 20 mice injected with B16-F10 cells developed metastases (median 10, range 0–60); 18 of 20 mice injected with B16-BL6 cells exhibited gross metastases (median 50, range 0–200). In all groups, well-isolated pulmonary metastases were surgically excised and established in culture as individual cell lines. The metastatic potential of cells from the parent tumour variants and several of their respective spontaneous metastases was determined in assays measuring experimental and/or spontaneous metastasis. A recent report suggested that the resection of a 'primary' B16 tumour could lead to the production of metastases[12], therefore we divided the mice bearing s.c. tumours into two treatment groups. In the first group, the leg carrying a 1.2–1.5 cm diameter tumour was amputated at mid-femur, thus including the popliteal lymph node, whereas in the second group of mice, the tumour was not resected.

Table 1 Ability of cells from B16 melanoma variants and their spontaneous metastases to form lung tumour colonies

Tumour line	No. of cells injected ($\times 10^3$)	Median no. of pulmonary foci (range)	P
B16-F1	50	5 (2–35)	—
M1	50	75 (30–>300)	0.003
M2	50	61 (29–103)	0.02
M3	50	91 (0–237)	0.002
M4	50	127 (41–>300)	0.0002
B16-BL6	50	48 (19–237)	—
M1	50	>300 (80–>300)	0.002
M2	50	170 (1–262)	0.03
M3	50	>300 (170–>300)	0.001
M4	50	>300 (172–>300)	0.0002
B16-F10	10	11 (1–19)	—
M1	10	11 (6–38)	0.86
M2	10	21 (3–42)	0.08
M3	10	6 (3–25)	0.47
M4	10	17 (4–35)	0.078
B16-F10	50	152 (22–219)	—
M1	50	217 (97–297)	0.12
M2	50	251 (1–300)	0.02
M3	50	130 (41–156)	0.3
M4	50	183 (93–>300)	0.96

Spontaneous metastases (M1–M4) from each of the tumours were obtained from mice bearing footpad tumours. Cells were collected from mid-log phase cultures by 1 min treatment with 0.25% trypsin, 0.02% EDTA solution. Single-cell suspensions of >95% viability were adjusted to the desired concentrations in Hanks' balanced salt solution ($Ca^{2+}+Mg^{2+}$-free) and 0.2 ml was injected into the lateral tail vein. Mice were necropsied 21 days later, and the number of lung tumour colonies was determined. Simultaneous comparisons of treatment versus control were based on the Kruskal–Wallis rank sums using the Miller approximation procedure[21]. For the analysis, >300 lung colonies were treated as being equal to 300. One-tailed P value is shown; 10 mice per group.

The data in Table 1 demonstrate that cells collected from spontaneous metastases from the poorly metastatic B16-F1 line produced significantly (15–25 times) more lung tumour colonies after i.v. implantation than an equal number of cells from the parent line. With the previously selected B16-BL6 line[16], cells collected from spontaneous metastases produced only 3–6 times more lung tumour colonies than the parent line. The B16-F10 cell line, which was specifically selected for the ability to produce pulmonary colonies[15], contrasted even more sharply with the B16-F1 cell line. Spontaneous metastases of B16-F10 tumours did not exhibit any increased propensity for experimental metastasis, regardless of whether the inoculum contained a high or low number of cells.

In the assays of spontaneous metastasis, individual variability in metastatic potential of cells populating different spontaneous metastases can be expected[2]. Thus, we analysed pooled data of metastases versus their respective parental tumours (Table 2). Once again, cells collected from spontaneous metastases originating from the B16-F1 tumour were significantly more metastatic than those originating from the parent tumour ($P = 0.007$). Cells collected from spontaneous metastases of B16-

Table 2 Production of spontaneous metastases by cells obtained from parent B16 melanoma variants and their spontaneous metastases

Tumour line	Incidence of spontaneous metastasis in mice with: Primary tumour resected	P	Primary tumour not resected	P	Total	P
B16-F1	3/7		2/10		5/17	
M1	4/7		6/10		10/17	
M2	7/7	0.025	7/11	0.009	14/18	0.007
M3	8/8		6/11		14/19	
M4	8/9		8/8		16/17	
B16-F10	6/8		2/8		8/16	
M1	8/10		5/10		13/20	
M2	9/10	0.34	7/10	0.05	16/20	0.0426
M3	8/9		7/10		15/19	
M4	9/10		6/9		15/19	
B16-BL6	8/8		6/8		14/16	
M1	8/8		4/8		12/16	
M3	10/10	0.64	5/10	0.31	15/20	0.2165
M3	6/6		6/8		12/14	
M4	6/8		7/12		13/20	

Primary tumours were initiated intramuscularly in 0.05 ml Hanks' balanced salt solution ($Ca^{2+}+Mg^{2+}$-free) into a posterior footpad. The tumour-bearing leg and popliteal lymph node were resected from 10 mice when the tumours reached 1.2–1.5 cm in diameter. Some mice died immediately after the resection and were thus excluded from the study. Morbid mice were necropsied, and the number of metastases was determined. The metastatic ability of the parent tumour was compared with that of cells recovered from metastases by the Fisher exact test using the data for resected or non-resected subcutaneous tumours as well as the combined data[21]. Combining the data from the two treatment groups increased sample size and yielded a more sensitive test.

F10 tumours exhibited an increased propensity for spontaneous metastasis ($P = 0.0426$) but, as shown in Table 1, not for experimental metastasis. Finally, cells from spontaneous metastases produced by B16-BL6, a tumour that exhibits a very high incidence of spontaneous metastasis (>80%), did not differ significantly ($P = 0.2165$) from the parent cells in metastatic potential.

Our findings demonstrate that metastases produced from the unselected, heterogeneous, and poorly metastatic B16-F1 cell line indeed contained cells having enhanced metastatic potential. In contrast, when the parent population was previously selected and was already performing at near peak ability (B16-F10 for lung colonization or B16-BL6 for spontaneous metastasis), further selection for the particular phenotype could not be achieved. These results clearly support the concept that metastasis is a selective process and agree with the recent demonstrations[13,17] that non-selective processes (adaptation) do not result in the outgrowth of highly metastatic cells.

We have shown that the B16 melanoma, which originated in 1954, rapidly changes its characteristics depending on the *in vivo* conditions in which it is maintained[18]. For this reason the B16 tumour maintained in one laboratory may not be exactly equivalent to a B16 tumour propagated in another laboratory. Specifically, Trope[19] proposed that the repeated trocar or fragment passage of heterogeneous tumours leads to loss of their diversity and imposition of uniformity on the transplanted tumours. Recent studies have substantiated this hypothesis and have demonstrated that repeated passages with tumour fragments appear only to exacerbate this artefact[20]. Clearly, to study selection during the process of metastasis, it is critical that the starting tumour be heterogeneous. It is difficult to envisage how successful selection could occur with a tumour composed of a uniform cell population. Furthermore, as numerous artefacts can alter the outcome of metastatic assays[14], absolute standardization of experimental conditions is necessary if the results of such assays are to be reproducible. In any event, we conclude from our studies that in controlled, experimental conditions, metastases produced by the heterogeneous B16-F1 melanoma cell line have a specialized subpopulation of cells with enhanced metastatic properties, supporting the conclusion that metastasis is a selective process.

This research was sponsored by the NCI under contract No. NO1-CO-75380 with Litton Bionetics, Inc.

Received 9 December 1981; accepted 6 April 1982.

1. Fidler, I. J. & Kripke, M. L. *Science* **197**, 893–895 (1977).
2. Poste, G. & Fidler, I. J. *Nature* **283**, 139–146 (1980).
3. Nicolson, G. L. & Winkelhake, J. L. *Nature* **255**, 230–232 (1975).
4. Briles, E. B. & Kornfeld, S. *J. natn. Cancer Inst.* **60**, 1217–1222 (1978).
5. Brunson, K. W., Beattie, G. & Nicolson, G. L. *Nature* **272**, 543–545 (1978).
6. Tao, T. W. & Burger, M. M. *Nature* **270**, 437–438 (1977).
7. Talmadge, J. E., Meyers, K. M., Prieur, D. J. & Starkey, J. R. *Nature* **284**, 622–624 (1980).
8. Raz, A. & Hart, I. R. *Br. J. Cancer* **42**, 331–341 (1980).
9. Giavazzi, R., Alessandri, G., Spreafico, F., Garattini, S. & Mantovani, A. *Br. J. Cancer* **42**, 464–472 (1980).
10. Mantovani, A., Giavazzi, R., Alessandri, G., Spreafico, F. & Garattini, S. *Eur. J. Cancer* **17**, 71–76 (1981).
11. Eccles, S. A., Heckford, S. E. & Alexander, P. *Br. J. Cancer* **42**, 252–259 (1980).
12. Stackpole, C. W. *Nature* **289**, 798–800 (1981).
13. Raz, A., Hanna, N. & Fidler, I. J. *J. natn. Cancer Inst.* **66**, 183–189 (1981).
14. Fidler, I. J. *Meth. Cancer Res.* **15**, 399–439 (1978).
15. Fidler, I. J. *Nature new Biol.* **242**, 148–149 (1973).
16. Hart, I. R. *Am. J. Path.* **97**, 587–600 (1979).
17. Nicolson, G. L. & Custead, S. E. *Science* **215**, 176–178 (1982).
18. Fidler, I. J. & Nicolson, G. L. in *Cancer Biology Reviews* (eds Marchalonis, J., Hanna, M. G. Jr & Fidler, I. J.) 171–234 (Marcel-Dekker, New York, 1981).
19. Trope, C. in *Design of Models for Testing Cancer Therapeutic Agents* (eds Fidler, I. J. & White, R. J.) 62–79 (Van Nostrand Reinhold, New York, 1981).
20. Fidler, I. J. & Hart, I. R. *Cancer Res.* **41**, 3266–3267 (1981).
21. Miller, R. G. Jr *Simultaneous Statistical Inference*, 6–10, 167 (McGraw-Hill, New York, 1966).

Selective killing of malignant cells in a leukaemic rat bone marrow using an antibody–ricin conjugate

P. E. Thorpe*, D. W. Mason†, A. N. F. Brown*, S. J. Simmonds†, W. C. J. Ross‡, A. J. Cumber‡ & J. A. Forrester‡

* Imperial Cancer Research Fund, PO Box 123, Lincoln's Inn Fields, London WC2A 3PX, UK
† MRC Cellular Immunology Unit, Sir William Dunn School of Pathology, University of Oxford, Oxford OX1 3RE, UK
‡ Chester Beatty Research Institute, Fulham Road, London SW3 6JB, UK

Graft-versus-host disease and marrow graft rejection are two major problems in the treatment of leukaemia with whole body irradiation, high-dose chemotherapy and bone marrow transplantation. Both these problems could be avoided if the patient's own bone marrow could be rendered aleukaemic *in vitro* and then reinjected into the patient after the radio-chemotherapy had been completed. Here we describe a model system in which lethally irradiated rats were injected with a mixture of bone marrow cells and leukaemic cells and in which the development of leukaemia in the recipient rats was prevented by incubating the cells, before injection, with a monoclonal antibody–ricin conjugate that was selectively toxic to the leukaemic cells. Nonspecific toxicity of the conjugate to haematopoietic stem cells was blocked by carrying out the incubation in the presence of lactose which competitively antagonized the binding of the ricin moiety of the conjugate to galactose residues on the cell surface[1–3].

The rat leukaemia used in these experiments is of T-cell origin and developed originally in a PVG rat that had received two injections of radioactive tungsten trioxide ($^{185}WO_3$) into its lymphoid tissue approximately 26 weeks earlier[4]. Injection of graded doses of leukaemic cells into fresh PVG hosts showed that a 1,000-fold reduction in cell dose delayed the onset of the leukaemia by 8–10 days, indicating a cell cycle time of a little less than 24 h. The injection of 100 leukaemic cells into syngeneic recipients was always fatal and of two rats given approximately 10 cells, one developed leukaemia and the other did not.

The derivation of the monoclonal antibody, W3/25, with which the ricin was coupled is given in ref. 5. It is an IgGl class antibody and reacts with high affinity[5] with rat thymocytes[6], T

0028-0836/82/240594-03$01.00

helper cells[7,8] and macrophages[9] but not with bone marrow haematopoietic stem cells (D.W.M., unpublished). Although the leukaemic cells express the W3/25 antigen (B. Roser, personal communication), flow cytofluorograph analysis indicates that the level of antigen expression varies widely between cells. Attempts to produce a more homogeneous population by *in vivo* cloning have been unsuccessful.

Initially, we attempted to prepare a specific cytotoxic agent for the leukaemic cells by coupling W3/25 antibody by a disulphide bond directly to the ribosome-damaging A-subunit of ricin[10]. This approach was used successfully by Krolick *et al.*[11], who recently described a conjugate of rabbit anti-immunoglobulin antibody and ricin A-chain which was capable of killing leukaemic cells in infiltrated murine bone marrow *in vitro* without destroying the haematopoietic stem cells. The W3/25–ricin A-chain conjugate, however, had almost no cytotoxic action on the leukaemic cells or on rat splenic T lymphocytes[10], even when used at concentrations which saturated the cell-surface antigens. The difference in the cytotoxic performance of the two conjugates typifies the variable effectiveness of A-chain conjugates prepared in several laboratories (reviewed in refs 10, 12, 13) and suggests that only certain types of cell-surface antigen can mediate the translocation of the A-chain into the cytosol. Therefore, for the present work, we linked intact ricin to W3/25 antibody, thereby retaining any properties of the B-chain which are necessary to effect A-chain penetration. The conjugate was prepared using the *N*-hydroxysuccinimidyl ester of chlorambucil as described previously[10] and was used in the presence of lactose, which antagonized the binding of the ricin B-chain moiety to galactose residues on cells lacking the W3/25 antigen.

The W3/25–ricin conjugate was an extremely effective inhibitor of protein synthesis by PVG leukaemic cells in tissue culture, as shown by their markedly reduced capacity to incorporate ^{3}H-leucine. Leucine uptake was reduced by 50% on treatment of the cells for 1 h at 37 °C with the conjugate at a concentration of 1.4 ng ml^{-1} ricin (2.2×10^{-11} M). Even in the absence of lactose the W3/25–ricin conjugate was about 10-fold more potent than native ricin, and about 100-fold more potent than a control conjugate made with another monoclonal antibody, MRC OX8, which does not bind to the leukaemic cells (Fig. 1*a*). W3/25 antibody alone, at concentrations as high as 10 μg ml^{-1}, did not affect the rate of protein synthesis of the cells. When 100 mM lactose was included in the cultures, it strongly inhibited the nonspecific toxic actions of native ricin and of MRC OX8–ricin but was virtually without effect on the toxicity of the W3/25–ricin conjugate (Fig. 1*b*).

The percentage of cells which survived treatment with the W3/25–ricin conjugate *in vitro* was estimated by comparing the time that elapsed between the injection of the treated cells into PVG recipients and the appearance of leukaemic cells in the peripheral blood with that observed when graded doses of untreated leukaemic cells were injected. Rats which received 10^7 leukaemic cells that had been treated with the conjugate at 2.1 μg ml^{-1} ricin (10 μg ml^{-1} IgG) for 1 h at 37 °C in 100 mM lactose and then washed twice in lactose-containing medium developed leukaemia 8–10 days later than the recipients of untreated cells, indicating that the conjugate had destroyed about 99.9% of the leukaemic cells. Neither the treatment of cells in the presence of lactose with W3/25 antibody alone at 20 μg ml^{-1} nor treatment with MRC OX8–ricin at 4.3 μg ml^{-1} ricin delayed the appearance of leukaemia in the recipients. The leukaemic cells which grew in the rats that had received cells treated with the W3/25–ricin conjugate were as heterogeneous in expression of the W3/25 antigen as the parent population, and their treatment with the conjugate before transfer to further recipient rats again resulted in the destruction of at least 99.9% of the cells. Thus, the leukaemic cells which survived exposure to the conjugate do not seem to be a minor subpopulation which is insensitive to the conjugate or lacking in the W3/25 antigen. It is possible, however, that the cells which survived were at a stage in their cell cycle at which the W3/25 antigen was only weakly expressed.

To discover whether leukaemic cells in bone marrow could be killed by the antibody–toxin conjugate without destroying the haematopoietic stem cell activity, a mixture of such cells was incubated with the conjugate and, after washing, injected into PVG hosts which had been sublethally irradiated with 650 rad. The bone marrow cells were obtained from rats that were congeneic with PVG but whose immunoglobulin light-chain genes were derived from the DA strain and were therefore of the *1-a* allotype rather than the *1-b* allotype of PVG rats. Survival of stem cells in the injected mixture was assessed from their ability to compete with the stem cells in the irradiated

Table 1 In mixtures of leukaemic cells and bone marrow cells the malignant cells are specifically killed by the antibody–ricin conjugate

Expt	Cells injected	Incubation conditions	No. of recipients	Day of appearance of leukaemia	% B-cell chimaerism
1	10^3 Leuk. + 10^7 BM	Med.	3	16, 17, 18	—
		Med. + conj.	3	>67, >67, >67	34.0, 37.5, 38.6
	10^4 Leuk. + 10^7 BM	Med.	3	16, 16, 16	—
		Med. + conj.	3	36, >67, >67	—, 30.6, 46.5
	10^7 BM only	Med.	3	—	61.4, 66.2, 67.8
2	10^3 Leuk. + 10^7 BM	Med.	3	19, 19, 20	—
		Med. + conj.	3	—*, —*, >73	28.7
	10^4 Leuk. + 10^7 BM	Med.	3	12, 16, 17	—
		Med. + conj.	3	23, >73, >73	—, 27.8, 29.5
	10^7 BM only	Med.	3	—	52.2, 54.3, 55.0

Med., medium: Dulbecco (A + B) (phosphate-buffered saline + $CaCl_2$ and $MgCl_2$) + 100 mM lactose. Med. + conj., medium + conjugate: as for 'Med.' except that W3/25–ricin conjugate containing 2.1 μg ml^{-1} ricin and 10 μg ml^{-1} antibody was added. The conjugate had a molecular weight of 350,000 and contained antibody and ricin in a molar ratio of 2.0 : 1, suggesting that it comprised two molecules of antibody and one of toxin. It was more effective at destroying leukaemic cells in bone marrow than a simple conjugate (molecular weight 210,000) containing one molecule each of antibody and toxin. Recipient PVG rats were matched for age and sex. Rats in expt 1 were given 650 rad ^{137}Cs irradiation before cell transfer. Rats in expt 2 received 1,000 rad. The day of appearance of leukaemia is defined as the day at which the white blood cell count reached 2×10^4 mm^{-3}. All rats that reached this level subsequently became severely leukaemic, with a white blood cell count $>10^5$ mm^{-3}, at which time they were killed by ether anaesthesia. B-cell chimaerism was determined 39 days after cell injections for expt 1 and on day 32 for expt 2. Cervical lymph node cells were assayed for B-cell chimaerism by comparing the percentage of cells labelled by PVG rat F(ab')$_2$ anti-rat *1-a* allotype antibody with that labelled by rabbit F(ab')$_2$ anti-rat Fab. Both antibodies were conjugated with fluorescein (a gift of Dr S. V. Hunt). Lymph node cells were incubated with each of the two reagents and histograms of cell fluorescence obtained by flow cytofluorography using a Becton–Dickinson FACS II cell sorter. Using lymph node cells from a PVG *1-a* homozygous rat as a positive control, 37.1% of cells were shown to react with the anti-*1-a* allotype reagent and 39.4% with the anti-Fab antibody. For lymph node cells from a PVG donor the corresponding values were 1.8 and 35.1%, respectively. Leuk., leukaemic cells; BM, bone marrow cells.

* Died within 21 days of irradiation with no signs of leukaemia.

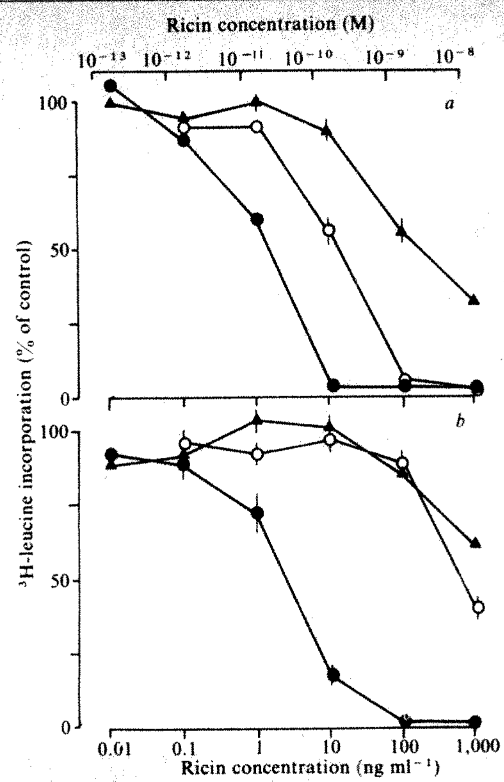


Fig. 1 Specific cytotoxic effects of ricin conjugated to monoclonal W3/25 antibody on PVG leukaemic cells in tissue culture. Blood was taken from a heavily leukaemic rat when its white blood cell count was 1.5×10^5 mm^{-3} and the leukaemic cells were separated by centrifuging the blood on Isopaque–Ficoll. The cells at a concentration of 2.5×10^6 ml^{-1} were incubated for 1 h at 37 °C with ricin (○), W3/25–ricin (●) or OX8–ricin (▲) in RPMI-1640 medium supplemented with 10% heat-inactivated fetal calf serum and antibiotics which either contained (*b*) or did not contain (*a*) 100 mM lactose. The cells were then washed in the same medium (but without toxin solutions) and incubated at 37 °C in fresh tissue culture medium without lactose. Their capacity to incorporate ^{3}H-leucine was measured 23 h later. Full details of the assay have been described previously[3]. The ^{3}H-leucine incorporation is expressed as a percentage of that in untreated control cultures which incorporated 10,300 c.p.m. per μCi. Each data point represents the geometric mean of three determinations, the standard deviations of which are indicated by vertical lines unless smaller than the points as plotted.

recipient and thus produce B lymphocytes that expressed the donor (*1-a*) allotype. The percentage chimaerism was compared with that obtained when untreated bone marrow cells, to which leukaemic cells had not been added, were injected into irradiated littermates of the recipients of the bone marrow/leukaemic cell mixtures. The evidence that the chimaerism assay measures stem cell activity is presented elsewhere[14]. The results of this experiment are presented in Table 1. All recipients of untreated bone marrow/leukaemic cell mixtures developed leukaemia before their level of B-cell chimaerism had been measured. None of the rats given W3/25–ricin-treated cell mixtures containing 10^3 leukaemic cells developed disease and of three recipients of inocula containing 10^4 leukaemic cells only one did so. All surviving rats developed a good level of chimaerism, although this was significantly lower in the recipients of bone marrow that had been exposed to the antibody–toxin conjugate than in control rats which had received bone marrow kept at 4 °C before injection. In a second experiment all recipients were lethally irradiated with 1,000 rad γ irradiation and survival of bone marrow stem cell activity was assayed, not by competition against surviving host bone marrow as in the first experiment but against a standard inoculum of 10^7 PVG bone marrow cells that were stored on ice until injection. As Table 1 shows, the results were very similar to those of the first experiment. Although acute radiation deaths limited the number of surviving animals in the group receiving 10^3 leukaemic cells, two out of three rats receiving 10^4 treated cells became long-term survivors with no sign of leukaemia and with significant chimaerism.

These experiments demonstrate that in the presence of 100 mM lactose the W3/25–ricin conjugate showed a high degree of target specificity. Approximately 99.9% of the leukaemic cells were killed by the conjugate while bone marrow stem cell activity was reduced by less than 50%. Some of the loss of stem cells can be attributed to effects other than ricin toxicity because incubation of the bone marrow in 100 mM lactose at 37 °C for 1 h in the absence of the antibody–ricin preparation resulted in a 13% reduction of chimaerism in assays similar to those described in Table 1.

The inclusion of lactose in the incubation medium not only prevented possible nonspecific killing of cells but had an important effect on the amount of conjugate bound nonspecifically to cells (such as erythrocytes) that was injected into the rats. Animals injected with 10^7 bone marrow cells that had been incubated for 1 h at 4 °C with the antibody–ricin conjugate at 2.1 μg ml^{-1} ricin in lactose-free medium and then washed twice died within 24 h of injection. Decreasing the concentration of conjugate to 0.21 μg ml^{-1} ricin prolonged the survival of recipient rats by only 24 h. However, when the cell incubations were carried out with conjugate in the presence of 100 mM lactose none of the recipient rats died or showed signs of ill health.

These results suggest that antibody–ricin conjugates might be used in conjunction with lactose to destroy malignant cells in human marrow transplants in situations where the antibody–A-chain conjugate is insufficiently effective.

Received 8 February; accepted 31 March 1982.

1. Olsnes, S., Saltvedt, E. & Pihl, A. *J. biol. Chem.* **249**, 803–810 (1974b).
2. Youle, R. J. & Neville, D. M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5483–5486 (1980).
3. Thorpe, P. E. *et al. Clin. exp. Immun.* **43**, 195–200 (1981).
4. Dibley, M., Dorsch, S. & Roser, B. *Pathology* **7**, 219–235 (1975).
5. Mason, D. W. & Williams, A. F. *Biochem. J.* **187**, 1–20 (1980).
6. Williams, A. F., Galfre, G. & Milstein, C. *Cell* **12**, 663–673 (1977).
7. White, R. A. H., Mason, D. W., Williams, A. F., Galfre, G. & Milstein, C. *J. exp. Med.* **148**, 664–673 (1978).
8. Brideau, R. J., Carter, P. B., McMaster, W. R., Mason, D. W. & Williams, A. F. *Eur. J. Immun.* **10**, 609–615 (1980).
9. Barclay, A. N. *Immunology* **42**, 593–600 (1981).
10. Thorpe, P. E. & Ross, W. C. J. *Immun. Rev.* **62**, 119–158 (1982).
11. Krolick, K. A., Uhr, J. W. & Vitetta, E. S. *Nature* **295**, 604–605 (1982).
12. Thorpe, P. E., Edwards, D. C., Ross, W. C. J. & Davies, A. J. S. in *Monoclonal Antibodies in Clinical Medicine* (eds Fabre, J. & McMichael, A.) (Academic, New York, in the press).
13. Olsnes, S. & Pihl, A. in *Pharmacology of Bacterial Toxins* (eds Drew, J. & Dorner, F.) (Pergamon, Oxford, in the press).
14. Hunt, S. V. & Fowler, M. H. *Cell Tissue Kinetics* **14**, 445–464 (1981).

Specificity of UV mutagenesis in the *lac* promoter of M13*lac* hybrid phage DNA

J. Eugene LeClerc & Nancy L. Istock

Department of Biochemistry, University of Rochester School of Medicine and Dentistry, Rochester, New York 14642, USA

The mutagenic effect of UV irradiation on *Escherichia coli* and its phages depends on induction of the *E. coli recA*$^+$ *lexA*$^+$ regulatory system (for review, see ref. 1). It is not understood how mutations result from UV damage in DNA, particularly with respect to the function(s) of proteins that may be induced in UV-irradiated cells. Misincorporation of nucleotides during DNA synthesis on template DNA containing UV-induced lesions may occur predominantly at the specific sites of damage, implying that the DNA replication apparatus in UV-irradiated cells can bypass premutational lesions. Alternatively, misincorporation may occur randomly, suggesting an altered and inaccurate DNA replication apparatus in UV-irradiated cells. To

study the specificity of UV mutagenesis at the nucleotide level, we have used M13*lac* hybrid phage, developed by Messing and colleagues[2,3] for rapid nucleotide sequencing of cloned DNA, as a forward mutation system. Nucleotide changes in a nonessential *lac* DNA sequence of the hybrid phage genome were determined by sequencing the DNA of mutant clones derived from UV-irradiated hybrid phage grown in UV-irradiated cells. We report here the nucleotide changes observed in the *lac* promoter of the mutant hybrid phage.

M13*lac* hybrid phage contains the M13 single-stranded DNA genome with an insert of the regulatory region and part (α-peptide) of the *lacZ* gene for lactose metabolism in *E. coli*. Infection of *E. coli* JM103 *lacZΔM15* cells with M13*lac* hybrid phage allows production of a functional β-galactosidase by intracistronic complementation. Infected cells are therefore lac^+, detected as blue plaques on indicator plates[2]. Inactivation of α-complementing activity, for example, by insertion of foreign DNA into the $lacZ'(\alpha)$ gene of the hybrid phage genome, gives rise to colourless plaques of infected cells, easily distinguishable from the blue plaques of M13lac^+-infected cells. In the studies reported here, inactivation of α-complementing activity was observed as a consequence of mutagenesis of UV-irradiated M13*lac* hybrid phage grown in UV-irradiated cells. Forward mutations in M13*lac* DNA were detected in the non-essential *lac* gene of the hybrid phage, so that all DNA changes (nucleotide substitutions, additions or deletions) resulting from UV mutagenesis could be observed by sequencing mutant clones.

Table 1 shows the mutation frequencies for induction of mutant *lac* hybrid phage derived from M13mp2 (ref. 3). Irradiation of M13mp2 hybrid phage at a UV dose giving ~0.01%

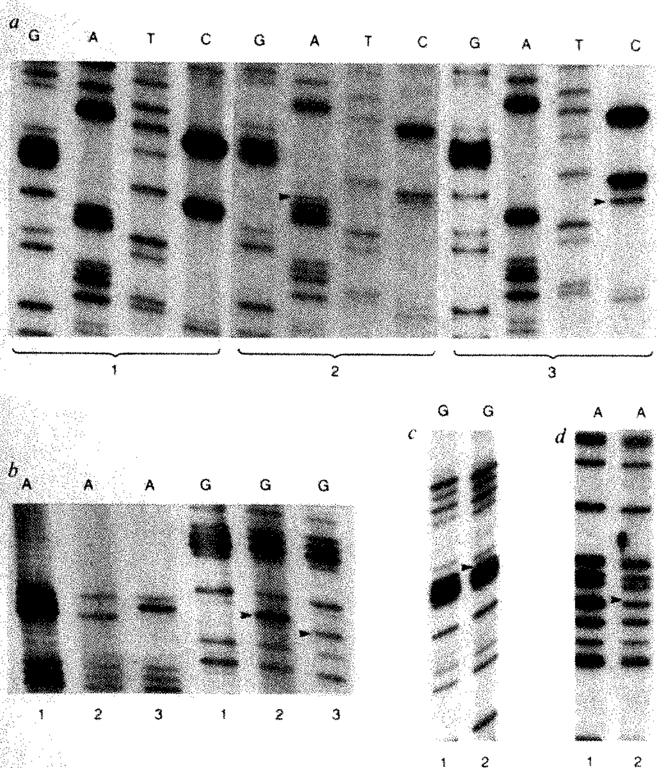


Fig. 1 Autoradiograms of nucleotide sequencing gels using templates prepared from M13mp2 hybrid phage and *lac* mutant derivatives. Preparation of templates, sequencing reactions and gel electrophoresis were carried out as described elsewhere[14] except that a 15-nucleotide primer (P.L. Biochemicals) was used. In each panel, lane 1 shows the sequence of normal M13mp2 DNA for comparison. The positions of nucleotide substitutions or deletions in mutant DNAs are indicated by arrowheads. *a*, C→T (viral strand) transition at position −37 (2) and T→G (viral strand) transversion at position −36 (3). *b*, T→C (viral strand) transitions at positions −35 (2) and −34 (3). *c*, C (viral strand) deletion (see text). *d*, T (viral strand) deletion (see text).

Table 1 Induction of *lac* mutations in M13mp2 hybrid phage

UV dose (J m^{-2}) Phage	Cells	Mutation frequency
0	0	$3.1\ (\pm1.6)\times10^{-4}$
125	0	$7.1\ (\pm1.3)\times10^{-4}$
0	50	$3.6\ (\pm5.1)\times10^{-4}$
125	50	$3.9\ (\pm0.2)\times10^{-3}$

E. coli JM103 were grown to 2×10^8 cells ml^{-1} at 37 °C in YT medium[13], then washed and resuspended in buffer (*p*H 8.0) consisting of 10 mM Tris, 67 mM KCl, 17 mM NaCl, 1 mM $MgSO_4$ and 1 mM $CaCl_2$. Cells were UV-irradiated at a dose rate of 0.8 J m^{-2} s^{-1} with stirring on ice. Cells were concentrated fivefold and resuspended in YT medium. Phage were diluted to 10^8 ml^{-1} in buffer and UV irradiation was carried out as for the cells. Cells were infected at multiplicity of ≤0.005 for 10 min at room temperature, plated onto indicator plates[2] and incubated overnight at 37 °C. All operations were carried out under yellow light. Mutation frequencies (±s.d.) are given as the ratio of plaques having mutant phenotypes to total plaques.

survival, or irradiation of *E. coli* JM103 host cells at a UV dose giving maximal UV reactivation of phage (N.L.I. and J.E.L., unpublished results) showed no statistically significant enhancement of mutation frequency over that for spontaneous production of mutant *lac* hybrid phage. UV irradiation of both phage and host cells caused an ~12-fold increase in mutation frequency. Direct plating of irradiated phage ensured that mutant clones were of independent origin. Mutant plaque phenotypes ranged from colourless to a medium blue that was detectably different from the blue plaques of M13mp2-infected cells on indicator plates, thus the colour reaction allowed detection of *lac* mutants completely or partially defective in α-complementing activity; 115 mutant clones resulting from UV irradiation of both phage and host cells were picked and purified for analysis of nucleotide changes.

The single-stranded DNA of mutant hybrid phages was sequenced using the chain termination method[4]; 40 of the 115 mutant DNAs sequenced showed nucleotide changes in the *lac* promoter of M13mp2 DNA. In Fig. 1, autoradiograms of nucleotide sequencing gels show examples of base substitutions and deletions identified in the *lac* promoter. Figure 2 shows a compilation of all sequence changes (viral strand) detected. Most of these mutations are clustered in three regions of the M13mp2 *lac* promoter: the −10 region or Pribnow box, the −35 region, and the binding site for the CAP protein. Single base substitutions account for over 80% of the changes induced by UV mutagenesis. Tandem base changes occurred at two sites and non-tandem base changes were found in one mutant clone (see Fig. 2 legend). Single base deletions occurred at three sites, although assignment of the exact positions of the deletions was impossible because they occurred at TT or CCCC sites. No nucleotide additions have yet been observed in UV-induced mutants.

Mutagenesis of the *lac* portion of M13mp2 hybrid phage is most enhanced by UV irradiation both of the host cells, to induce the cellular mutagenesis system, and of the phage, containing the target DNA molecule (Table 1). As pyrimidine dimers are the major and potentially lethal photoproducts induced by UV irradiation of DNA[5,6] the nucleotide changes resulting from UV mutagenesis of M13*lac* hybrid phage were analysed with respect to sites of potential pyrimidine dimer formation. Of all changes observed, 77% occurred at such sites, although most mutations were found in the −35 region where promoter mutations are expected (ref. 7 and references therein) and which is pyrimidine-rich. These sites do represent hotspots for UV mutagenesis, however, because preliminary analysis of collections of *lac* mutants induced by chemical mutagens has revealed additional nucleotide changes at positions −31, −45, −55 and −56 that were not observed as a consequence of UV irradiation (data not shown). All three nucleotide deletions occurred at adjacent pyrimidines. For base substitution

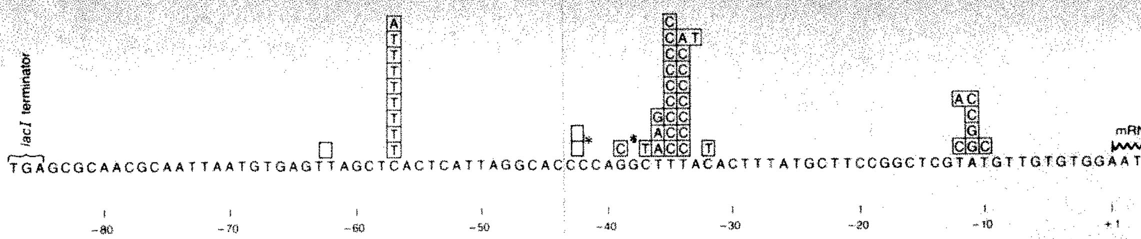


Fig. 2 Nucleotide changes in the *lac* promoter of mutant M13mp2 hybrid phage. The *lac* promoter sequence in hybrid phage DNA is numbered from the RNA polymerase initiation site. Each box represents a nucleotide change detected in *lac* mutant clones; nucleotide substitutions are shown in the box and nucleotide deletions are represented by open boxes. Tandem nucleotide substitutions are shown at positions −11, −12, and −34, −35. The exact positions of nucleotide deletions cannot be determined (see text). * Non-tandem nucleotide changes found in one mutant clone.

mutations at potential pyrimidine dimer sites, transitions were favoured over transversions (5:1), in agreement with previous studies on the specificity of UV mutagenesis in *E. coli*[8] and yeast[9]. It is particularly interesting that no tandem base substitutions occurred at potential pyrimidine dimer sites; the small proportion observed were at TA sites. Random substitution of nucleotides at pyrimidine dimer sites should yield up to 56% of changes at adjacent pyrimidines as tandem double changes, although the bias for transitions in UV mutagenesis should decrease that frequency.

Brandenburger *et al.*[10] determined the DNA sequences of revertants of M13 amber mutant phages grown in conditions of UV irradiation similar to those reported here. They concluded that pyrimidine dimers are not the main sites of base substitution mutagenesis in UV-irradiated M13 DNA. Unlike the M13*lac* forward mutation system, the greatest enhancement of mutations in the M13 reversion system is observed by UV-irradiating the host cells alone, so that nucleotide changes would not be expected to occur at potential pyrimidine dimer sites. In the case of UV irradiation of M13*am7H3* grown in UV-irradiated bacteria, however, Brandenburger *et al.*[10] did observe a fourfold increase in reversion frequency over that induced by UV irradiation of host cells alone, and the sequencing data showed a spectrum of mutations distinctly shifted to nucleotide changes at the 3′ pyrimidine of adjacent thymines.

Clearly, mutagenesis in UV-irradiated cells occurs at sites not directly associated with pyrimidine dimers. The principal sites of misincorporation may be at pyrimidine dimers, however, in which case these lesions in DNA templates are not totally non-instructive (compare ref. 8). The paucity of UV-induced tandem base changes at adjacent pyrimidines in mutant M13*lac* hybrid phage, M13 phage[10], *E. coli*[11] and yeast[12] supports this view.

We thank Diana Bongartz for technical assistance and Chris Lawrence for helpful discussions. This work was supported by a Cottrell Research Grant from Research Corporation, an ACS Institutional Research Grant, and NIH grant GM 27817.

Received 4 January; accepted 5 April 1982.

1. Witkin, E. M. *Bact. Rev.* **40**, 869–907 (1976).
2. Messing, J., Gronenborn, B., Müller-Hill, B. & Hofschneider, P. H. *Proc. natn. Acad. Sci. U.S.A.* **74**, 3642–3646 (1977).
3. Gronenborn, B. & Messing, J. *Nature* **272**, 375–377 (1978).
4. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
5. Setlow, J. K. & Setlow, R. B. *Nature* **197**, 560–562 (1963).
6. Setlow, R. B. & Setlow, J. K. *Proc. natn. Acad. Sci. U.S.A.* **48**, 1250–1257 (1962).
7. Siebenlist, U., Simpson, R. B. & Gilbert, W. *Cell* **20**, 269–281 (1980).
8. Lawrence, C. W. *Molec. gen. Genet.* **182**, 511–513 (1981).
9. Lawrence, C. W. & Christensen, R. B. *Molec. gen. Genet.* **177**, 31–38 (1979).
10. Brandenburger, A. *et al. Nature* **294**, 180–182 (1981).
11. Coulondre, C. & Miller, J. H. *J. molec. Biol.* **117**, 577–606 (1977).
12. Lawrence, C. W., Stewart, J. W., Sherman, F. & Christensen, R. *J. molec. Biol.* **85**, 137–162 (1974).
13. Miller, J. H. *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, New York, 1972).
14. Heidecker, G., Messing, J. & Gronenborn, B. *Gene* **10**, 69–73 (1980).

Expression of human β-interferon cDNA under the control of a thymidine kinase promoter from herpes simplex virus

Gregory R. Reyes*, Elizabeth R. Gavis*, Alexander Buchan*, N. Babu K. Raj†, Gary S. Hayward* & Paula M. Pitha†

* Department of Pharmacology and Experimental Therapeutics and † Oncology Center, Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, Maryland 21205, USA

The availability of an isolated interferon gene cDNA clone and biological assays for its expression, provide an opportunity to examine the functional role of sequences preceding the coding portion of the interferon gene. In the present study, we examined whether the human β-interferon (IFN-β) gene can be transcribed and induced efficiently after replacing the normal 5′ flanking sequences[1,2] with a defined promoter segment from another gene. For this purpose the coding portion of the IFN-β gene[3] was inserted downstream from the promoter sequence and RNA start site of the thymidine kinase (TK) gene from herpes simplex virus type 1 (HSV-1; ref. 4). Expression of interferon activity under the control of the viral TK promoter in this chimaeric plasmid was demonstrated by microinjection into the nuclei of *Xenopus* oocytes[5,6] and by transfection into mouse cells[7] followed by superinfection with herpes simplex virus.

The interferon gene is particularly suitable for functional analyses because the high specific activity of the interferon protein product permits detection of picogram quantities in a sensitive biological assay[8], and its species specificity allows identification of the interferon synthesized from the added heterologous DNA as opposed to that encoded by endogenous host cell genes. Our objectives were both to insert the 560-base pair (bp) structural gene sequence of the pHFβ-cDNA clone[9] downstream from the HSV-1 TK promoter sequence, and to have a second selectable intact TK marker present to facilitate the subsequent introduction of the hybrid TK–IFN cDNA gene into Ltk⁻ cells. The construction of plasmids containing the TK promoter fused to the coding sequences of interferon cDNA in both the sense (pGR192) and nonsense (pGR191) directions is summarized in Fig. 1, together with the principal features of the parent pGR18, pGR156 and pHFβ-cDNA plasmids. In pGR192, the coding (or sense) strand of the interferon insert lies downstream from the HSV-1 TK promoter, 54 nucleotides after the 5′ transcription initiation site but in front of the first AUG for the TK protein[4,10].

To determine whether the HSV TK flanking sequences can promote valid transcription of the IFN-β gene, plasmids containing the hybrid TK–IFN cDNA gene were tested for the ability to direct synthesis of interferon after microinjection into

nuclei of *Xenopus* oocytes. We compared the levels of interferon synthesized in oocytes injected with DNA from the pHFβ-cDNA, pGR191 and pGR192 clones with those from oocytes injected with pBR322 and pGR18 (Table 1). In both experiments, oocytes receiving the pGR192 clone produced interferon (120–530 U ml^{-1}), whereas no interferon activity

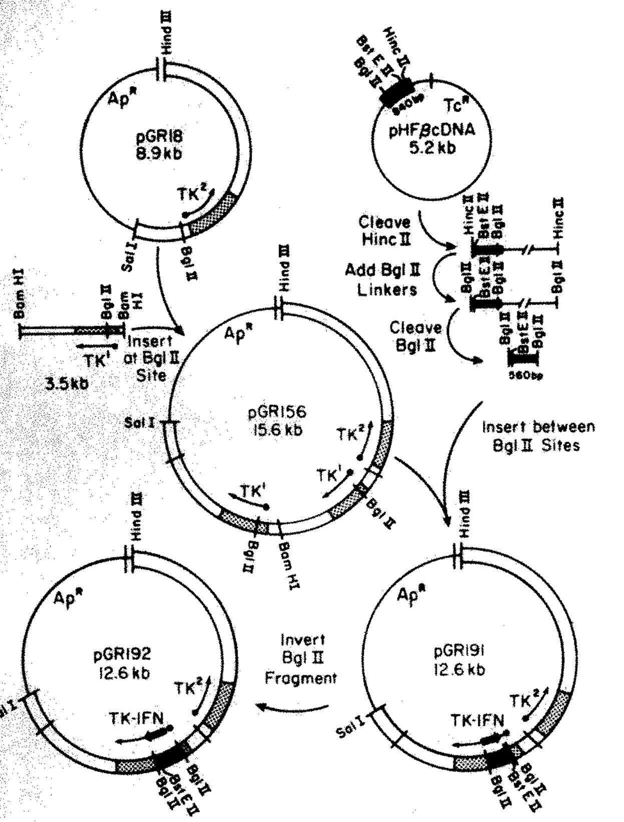


Fig. 1 Diagram illustrating the general strategy used to construct an interferon expression plasmid containing the coding portion from cloned IFN-β cDNA linked to the TK gene. The pGR18 plasmid[26] contains a cloned 4.8-kb *Hin*dIII–*Sal*I fragment from HSV-2(333) viral DNA (open bars), which includes the HSV-2 TK gene (stippled bars), inserted into pBR322 (single line). The 3.5-kb *Bam*H1–O fragment shown contains the HSV-1 TK gene and was derived originally from the HSV-1(MP) strain[6]. Arrows indicate the direction of transcription in each of the TK genes and solid circles represent the promoter regions. The pHFβ-cDNA clone described previously[9] was derived from a cDNA library of human fibroblast cells induced with poly(rI·rC) and was inserted into the *Pst*I site of pBR322. The 840-bp human fibroblast cDNA segment of this plasmid (solid bars) possesses single *Bgl*II, *Bst*EII, *Hin*cII, *Pst*I, *Pvu*II and *Taq*I cleavage sites in the same arrangement as those reported by Derynck *et al.*[3] for their 860-bp IFN-β cDNA plasmid. We have used only the 560-bp *Hin*cII–*Bgl*II fragment encoding the IFN-β structural gene which includes the AUG translation initiation codon for the signal peptide and the UGA termination signal. To appropriately modify the 560-bp interferon fragment phosphorylated *Bgl*II linkers[27] were ligated at 20-fold molar excess to the *Hin*cII-cleaved pHFβ-cDNA plasmid at 10 °C for 30 h in a reaction volume of 50 μl (refs 28, 29). Gel electrophoresis of a series of time point samples confirmed that linkers were ligated to the *Hin*cII fragments, and that formation of higher multimers of the *Hin*cII fragments with and without linker attachment had also occurred. After ligation, the DNA was ethanol-precipitated and resuspended in 0.01 M Tris-HCl, 0.001 M EDTA *p*H 8.0. Appropriate amounts of $MgCl_2$ and NaCl were added in order for *Bgl*II digestion to generate cohesive extensions at the 5′ end of the interferon cDNA and also to cleave at the *Bgl*II site at the 3′ end of the coding sequence. This fragment was then inserted between the two *Bgl*II sites in the two tandemly repeated HSV-1 TK genes within pGR156. Cultures of *Escherichia coli* HB101 were transformed by the calcium shock method of Mandel and Higa[30]. For hybridization screening, the transfection mixture was spread on to selective ampicillin plates and grown until the colonies reached a size of 0.1–0.2 mm. The colonies were then picked up on the surface of keyed Whatman No. 1 filter paper circles and transferred (colonies uppermost) to chloramphenicol plates for further incubation (12–48 h). The filters were treated with alkali, neutralized, washed and baked essentially as described by Hanahan and Meselson[31]. Positive clones on these filters were detected directly by hybridization with appropriate nick-translated probes. Digestion of selected plasmids containing a single 560-bp *Bgl*II insert with *Pst*I, *Pvu*II or *Bst*EII, followed by hybridization to Southern transfers from agarose gels, revealed that the interferon sequence in pGR191 was oriented in the opposite direction to that of the TK promoter, whereas that in pGR192 represented the correct orientation.

(<20 U ml^{-1}) could be detected with any of the other plasmids. The levels of interferon obtained were comparable to those synthesized in oocytes after microinjection with poly(A)-containing RNA isolated from human fibroblasts after induction with poly(rI·rC)[11,12]. These results indicate that efficient synthesis of interferon mRNA occurred only when the cDNA coding sequence was inserted in the correct orientation downstream from the TK promoter.

The presence of an intact HSV-2 TK gene in our expression plasmids was utilized to establish permanent TK^+ cell lines by calcium transfection of Ltk^- cells and selection of colonies growing in hypoxanthine–aminopterin–thymidine (HAT) medium[13]. Five TK^+ clones from the pGR191 transfections and five from the pGR192 transfections were tested, but none produced interferon constitutively (detection limit 1 U ml^{-1}). As the expression of an active resident viral TK gene in mouse L-cell lines has been shown to be enhanced 5–10-fold by superinfection with HSV-1 (tk^-) virus[14], we reasoned that the HSV-1 TK promoter attached to the interferon cDNA, although apparently inactive for constitutive synthesis in our cell lines, might nevertheless respond to the viral regulatory functions introduced by a superinfecting viral genome. Table 2 shows that the LH_2p192–8 subclone does indeed express interferon activity induced by HSV-1 and HSV-2, but not by infection with more distantly related viruses of the herpes group, for example, cytomegalovirus (CMV) or pseudorabies virus (PRV).

The interferon synthesized both in microinjected oocytes and LH_2p192–8 cells displayed the same species specificity and antigenicity as that derived from human fibroblasts, that is, it was active on human cells but not on bovine or mouse cells, and it was completely neutralized by anti-human IFN-β antiserum[15], but not by anti-human α-interferon[16] or anti-mouse IFN-α plus -β antisera[17]. We conclude that the interferon synthesized is authentic human IFN-β, which could only have been encoded by the exogenous human IFN-β sequences introduced into these mouse cells by transfection.

The kinetics of IFN-β synthesis in superinfected LH_2p192–8 cells, including initiation within 2–3 h, maximal expression at 6 h and a decrease by 20 h, follows the expected time course for viral transcriptional regulation of a delayed-early promoter, that is, induction by immediate-early functions followed by repression after synthesis of late functions. Additional experiments with a viral *ts* mutant that synthesizes an inactive immediate-early protein supported this conclusion (Table 2, expt 2). The *ts*B2 mutation maps in the HSV-1 IE gene IV, which encodes a phosphorylated protein of molecular weight 175,000 (175K) having DNA binding properties and a nuclear chromatin-associated location[18,19]. This protein has been shown

Table 1 Human IFN-β synthesized in *Xenopus* oocytes

Plasmid injected	DNA ($\mu g\,\mu l^{-1}$)		Interferon (U ml^{-1}) GM2504		Ly	
	a	*b*	*a*	*b*	*a*	*b*
pHFβ-DNA	0.20,	0.10	<20,	<20	<20,	—
pGR192	0.10,	0.10	530,	120	<20,	—
pGR191	0.17,	0.10	<20,	<20	<20,	—
pBR322	0.14,	0.10	<20,	<20	<20,	—
pGR18	0.15,	—	<20,	—	<20,	—

Mature *Xenopus* oocytes were prepared as described by McKnight and Gavis[6] and a sample (3–5 ng) of supercoiled DNA in a 30-nl volume was microinjected into each nucleus. Fifty oocytes per assay were washed and incubated in 3 ml of Barth's medium at 18 °C for 36 h. Interferon activity in the medium was assayed as described previously[11,12] using the cytopathic assay of Finter [24] on either human fibroblast cells trisomic for chromosome 21 (GM2504) or on mouse Ly cells. The interferon titres are given in international reference standard units (69/17). One unit of standard human IFN-β titrated as 1 unit in our assay on GM2504 cells and as 10^{-3} units on Ly cells. The results of two independent experiments are shown: *a*, *Xenopus borealis* oocytes; *b*, *Xenopus laevis* oocytes. —, Not determined.

to have a direct role, during HSV infection, in the transcriptional activation (or depression) of viral delayed-early genes including the thymidine kinase gene[20,21]. Infection of LH_2p192–8 cells with HSV-1 (*ts*B2) in non-permissive conditions (39.5 °C) resulted in a 50-fold reduction in the level of IFN-β synthesis compared with infection by the same virus in permissive conditions (33.5 °C).

Induction of interferon activity was not detected after HSV-1 infection in five LH_2p191 cell lines and in four other LH_2p192 cell lines when tested at sixth passage (only one of each is shown in Table 2). Surprisingly, in earlier experiments with these cell lines at third passage, we found that human IFN-β could be induced in eight of the ten LH_2p191 or LH_2p192 clones by either poly(rI·rC) treatment or Newcastle disease virus infection, indicating that the IFN-β cDNA sequences were present and that they also respond to induction by some other mechanism that does not require the presence of the TK promoter[13]. The inability to induce interferon in some LH_2p192 lines after HSV infection could represent either sequence rearrangements or loss of the TK–IFN hybrid gene after integration. To examine this question, DNA samples from LH_2p191 and LH_2p192 cell lines were digested with *Pvu*II (which cleaves once within the IFN-β structural sequence) and hybridized with a ^{32}P-labelled fragment probe consisting of the isolated 560-bp IFN-β coding sequence (Fig. 2). The single-copy IFN-β gene sequences present in human fibroblast cell DNA gave two bands of the size expected from previous reports[22,23]. Similarly, the LH_2p192–8 and LS_2p191-1 cell DNAs each exhibited two bands that correlated with the appropriate fragment sizes expected after cleavage at the flanking *Pvu*II sites in the HSV-2 TK gene. In contrast, DNA from the LH_2p192-1 line at sixth passage did not hybridize with the interferon cDNA probe. Therefore, the lack of inducibility observed in the LH_2p192-1 line correlated with the absence of IFN-β sequences from the cellular DNA, whereas the lack of induction in the LH_2p191-3 line was probably related to the reverse orientation of the promoter with respect to the IFN-β coding sequence. These data strongly support the notion that the synthesis of IFN-β in the LH_2p192-8 transfected mouse cell line represents transcriptional events related directly to the correct positioning of the TK promoter sequence.

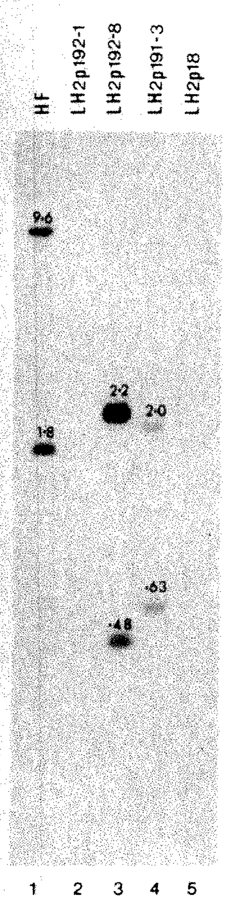


Fig. 2 Demonstration of the presence of human IFN-β DNA in transfected mouse cells. High molecular weight DNA was extracted from mass cultures of appropriate subcloned cell lines at sixth passage, digested to completion with *Pvu*II and fractionated by agarose gel electrophoresis (1.2% horizontal gels, 10 µg DNA per sample). The DNA was transferred to nitrocellulose and incubated with a ^{32}P-labelled probe prepared by nick-translation of the isolated 560-bp IFN-β fragment from pGR191. Hybridization was carried out at 37 °C for 18 h in 50% formamide, 10% dextran sulphate, 1×Denhardt's mix, 0.5% SDS, 100 µg ml^{-1} denatured salmon sperm DNA, 20 µg ml^{-1} *E. coli* RNA, 3×SSC, 1 mM EDTA, and 50 mM Tris-HCl *p*H 7.6, by a procedure based on that described elsewhere[32]. The sizes of hybrid bands detected in the autoradiograph are given in kilobase pairs. Lane 1, human fibroblast HF-4 cell DNA used as a single-copy reference; lane 2, LH_2p192-1 DNA; 3, LH_2p192-8 DNA; 4, LH_2p191-3 DNA; 5, LH_2p18 DNA.

Table 2 Induction of human IFN-β in transfected mouse cells

		Interferon induced (U ml^{-1})			
			Superinfection†		
Cell line	Virus used	Not infected*	3 h	6 h	20 h
Expt 1					
Ltk$^-$	HSV-1(MP), 36 °C	<5	<5	<5	<5
LH_2p18	HSV-1(MP), 36 °C	<5	<5	<5	<5
LH_2p191-3	HSV-1(MP), 36 °C	<5	<5	<5	<5
LH_2p192-1	HSV-1(MP), 36 °C	<5	<5	<5	<5
LH_2p192-8	HSV-1(MP), 36 °C	<5	60	160	45
Expt 2					
Ltk$^-$	HSV-1(MP), 36 °C	<5	—	<5	—
LH_2p192-8	HSV-1(MP), 36 °C	<5	—	320	80
LH_2p192-8	HSV-1(KOS*ts*B_2), 33.5 °C	—	—	720	40
LH_2p192-8	HSV-1(KOS*ts*B_2), 39.5 °C	—	—	15	<5
LH_2p192-8	HSV-2(333), 36 °C	—	—	720	—
LH_2p192-8	CMV(Towne), 36 °C	—	—	<5	<5
LH_2p192-8	PRV, 36 °C	—	—	<5	—
HF-4	HSV-1(MP), 36 °C	<5	—	<5	—

The transfection protocol used was essentially that described by Graham *et al.*[25] as modified for TK selection by Wigler *et al.*[7] and is described in detail elsewhere[13]. Ltk$^-$ cells are thymidine kinase-negative L cells; LH_2p18, LH_2p191 and LH_2p192 cells are TK$^+$ clones selected in HAT medium after transfection of Ltk$^-$ cells with the pGR18, pGR191 and pGR192 plasmids, respectively. HF-4 cells are diploid human fibroblasts.

* Media from 5×10^5 cells per 35 mm dish were examined directly for the constitutive production of interferon (24 h collection) on human GM2504 indicator cells.

† Cell cultures were superinfected with wild-type herpesviruses at 36 °C or with HSV-1(KOS*ts*B_2) at either 33.5 °C (permissive conditions) or 39.5 °C (non-permissive conditions). The multiplicities of infection ranged from 10 plaque forming units per cell in expt 1 to 30 plaque forming units per cell in expt 2. The levels of interferon accumulated in the medium at various times after virus absorption were assayed on GM2504 cells.

By placing the interferon gene behind the TK promoter, we hoped that an active product might be expressed continuously in transfected cells in the same way that the TK gene product is expressed continuously in TK$^+$ cell lines. However, this was not the case. Although we selected for constitutive expression of the viral thymidine kinase enzyme behind one TK promoter in the LH_2p192-8 cell line, a second TK promoter introduced into the same cells, but controlling the unselected interferon cDNA gene, was apparently not expressed to any significant degree. Nevertheless, this second TK promoter was still available for transcriptional activation when the appropriate viral factors entered the cell. Therefore, the control of expression from the resident unselected TK promoter in the transfected mouse cells follows the normal pattern seen in HSV-infected cells. The fact that HSV infection does not induce interferon in a human fibroblast cell line (HF-4), and the lack of induction by the HSV-1 *ts*B2 mutant in LH_2p192-8 cells at non-permissive temperature, provided further evidence that it was neither the input virions nor viral DNA which activated the TK–IFN hybrid gene, but rather a specific HSV regulatory gene product (probably the 175K protein itself).

We thank Dr W. Fiers for the gift of the HF12cDNA clone used in the original identification of our pHFβ-cDNA plasmid, Drs K. Berg and E. DeMaeyer and J. DeMaeyer for gifts of antisera, and Dr B. Sollner-Webb for making available microinjection facilities. G.R.R. was a predoctoral fellow in the Medical Scientist Training Program at the Johns Hopkins School of Medicine (GM 07309). A.B. was supported by a NATO exchange program fellowship. This work was funded by NIH grants 5ROI AI-10944-09 (P.M.P.) and 2ROI CA-22130-04 (G.S.H.).

Received 5 January; accepted 26 March 1982.

1. Nagata, S., Mantei, N. & Weissman, C. *Nature* **287**, 401–408 (1980).
2. DeGrave, W., Derynck, R., Tavernier, J., Haegeman, G. & Fiers, W. *Gene* **14**, 137–143 (1981).
3. Derynck, R. *et al. Nature* **285**, 542–547 (1980).

4. McKnight, S. L. *Nucleic Acids Res.* **8**, 5949–5964 (1980).
5. Mertz, J. E. & Gurdon, J. B. *Proc. natn. Acad. Sci. U.S.A.* **74**, 1502–1506 (1977).
6. McKnight, S. L. & Gavis, E. R. *Nucleic Acids Res.* **8**, 5931–5948 (1980).
7. Wigler, M. *et al. Cell* **11**, 223–232 (1977).
8. Knight, E. Jr *J. Interferon Res.* **2**, 1–12 (1980).
9. Raj, N. B. K. & Pitha, P. M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 7426–7430 (1981).
10. McKnight, S., Gavis, E. R., Kingsbury, R. & Axel, R. *Cell* **25**, 385–398 (1981).
11. Reynolds, F. H. Jr, Premkumar, E. & Pitha, P. M. *Proc. natn. Acad. Sci. U.S.A.* **72**, 4881–4885 (1975).
12. Raj, N. B. K. & Pitha, P. M. *Proc. natn. Acad. Sci. U.S.A.* **74**, 1483–1487 (1976).
13. Pitha, P. M. *et al. Proc. natn. Acad. Sci. U.S.A.* (in the press).
14. Leiden, J. M., Buttyan, R. & Spear, P. *J. Virol.* **20**, 413–424 (1976).
15. Neurath, A. R., Strick, N., Raj, N. B. K. & Pitha, P. M. *J. Interferon Res.* **2**, 51–57 (1982).
16. Berg, K. *Scand. J. Immun.* **114**, 77–86 (1977).
17. DeMaeyer-Guignard, J., Tovey, M. G., Gresser, I. & DeMaeyer, E. *Nature* **271**, 622–625 (1978).
18. Cabral, G. A., Courtney, R. J., Schaffer, P. A. & Marciano-Cabral, F. *J. Virol.* **33**, 1192–1198 (1980).
19. Hay, R. T. & Hay, J. *Virology* **104**, 230–234 (1980).
20. Preston, C. M. *J. Virol.* **29**, 275–284 (1979).
21. Watson, R. J. & Clements, J. B. *Nature* **285**, 329–330 (1980).
22. Tavernier, J., Derynck, R. & Fiers, W. *Nucleic Acids Res.* **9**, 461–471 (1981).
23. Houghton, M. *et al. Nucleic Acids Res.* **9**, 247–266 (1981).
24. Finter, N. B. *J. gen. Virol.* **5**, 419–427 (1969).
25. Graham, F. L., Veldhuisen, G. & Wilkie, N. M. *Nature* **245**, 265–266 (1973).
26. Hayward, G. S., Reyes, G. R., Gavis, E. R. & McKnight, S. L. in *Herpesvirus DNA. Developments in Molecular Virology* Vol. 1 (ed. Becker, Y.) 271–306 (Martinus-Nijhoff, The Hague, 1981).
27. Sakonju, S., Bogenhagen, D. F. & Brown, D. D. *Cell* **19**, 13–25 (1980).
28. Dugaiczyk, A., Boyer, H. W. & Goodman, H. M. *J. molec. Biol.* **95**, 171–184 (1975).
29. Lillehaug, J. R., Kleppe, R. K. & Kleppe, K. *Biochemistry* **15**, 1858–1865 (1976).
30. Mandel, M. & Higa, A. *J. molec. Biol.* **53**, 159–162 (1970).
31. Hanahan, D. & Meselson, M. *Gene* **10**, 63–67 (1980).
32. Sager, R., Anisowicz, A. & Howell, N. *Cell* **23**, 41–50 (1981).

Unique insertion site of Tn*7* in the *E. coli* chromosome

Conrad Lichtenstein* & Sydney Brenner

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Among transposable elements, transposon Tn*7* is unusual in that it inserts into a particular site of the *Escherichia coli* chromosome with high efficiency[1,2] and with a unique orientation[2]. Most other transposable elements in prokaryotes are capable of translocating from and to a variety of sites in bacteriophage and plasmid DNA as well as in the chromosome[3]. We have previously shown that the 'attachment site' for Tn*7* in *E. coli* is located between *oriC* and *glmS* at minute 82 (ref. 4) of the chromosome[2]. We now report the DNA sequences of the termini of Tn*7*, of the attachment site (before and after Tn*7* insertion) and of an alternative site of Tn*7* attachment. We conclude that the specificity of the attachment site is determined by specific DNA sequences and that secondary structures may be involved.

Tn*7* is a very large (14 kilobase (kb)) element which codes for resistance to the antibiotics trimethoprim, streptomycin and spectinomycin[1]. Its preference for the attachment site is so pronounced that, when the site is available, virtually all transposition events are directed there[2]. It is not yet known whether the high efficiency with which Tn*7* transposes into *Pseudomonas aeruginosa*[5], *Klebsiella pneumoniae* and *Agrobacterium tumefaciens* (M. van Montagu, personal communication) involves a similarly unique site, but the transposition of Tn*554* in *Staphylococcus aureus* is known to involve such a site[6].

We have obtained the DNA sequence of the attachment site by the use of the plasmid pCPL6[2], previously constructed from the plasmid pBR322[7] by replacing the small *EcoRI–Bam*HI restriction fragment with the much larger (8×10^6 molecular weight) corresponding fragment derived from the transducing phage[8] λ *asn*5, which was shown to contain the Tn*7* attachment site[2]. The site-specific properties are retained[2]. Figure 1 shows a restriction map of pCPL6 including the Tn*7* insertion-site. For the DNA sequence of the non-preferred site, we have used the plasmid Col E1: :Tn*7*. In both cases, DNA sequences of appropriate restriction fragments were subcloned into an M13 phage vector and sequenced using the chain termination method[9].

*Present address: Department of Microbiology and Immunology, University of Washington, Seattle, Washington 98195, USA.

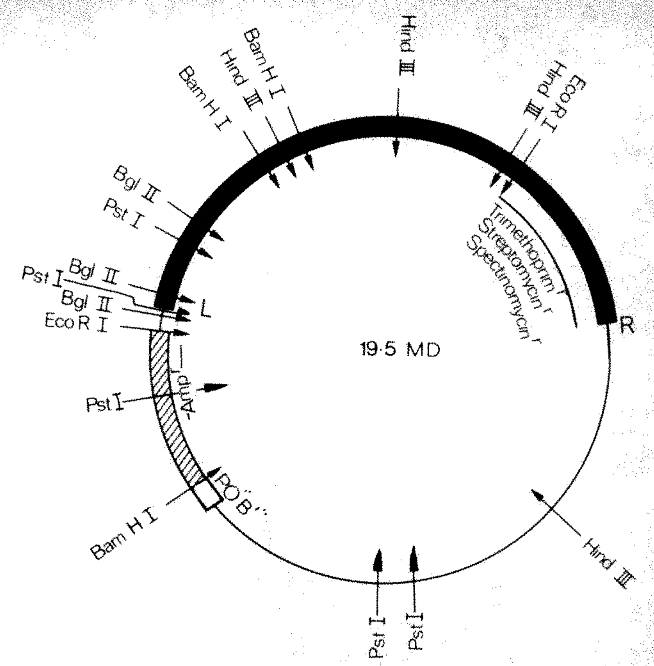


Fig. 1 Restriction map of pCPL6: :Tn*7*, showing the source of the sequenced material. A small *Bgl*II junction fragment at the left end (L) of Tn*7* was subcloned into the phage vector M13mp2/Bam (ref. 18). This fragment was used as a probe to identify, by plaque hybridization screening[19], the *Sau*3A restriction fragment from pCPL6 that includes the Tn*7* attachment site before insertion. The cloned insert of this latter phage was used similarly as a probe to identify phage re-combinants containing the *Sau*3A junction fragment of the right end (R) of Tn*7* in pCPL6: : Tn*7*. This right-end junction fragment was then used as a probe to identify the *Sau*3a junction fragment at the right end of Tn*7* in the plasmid Col E1: : Tn*7*. In addition, a *Pst*I fragment overlapping with the left-end *Bgl*II junction fragment in pCPL6: : Tn*7* was subcloned into the M13mp2/Pst phage vector[20]. The cloned inserts of all these M13 phage recombinants were sequenced by the chain terminator method[9] using the virion DNA as template[21] and a 17-bp synthetic primer (M. Gait, unpublished) or the 92-bp universal primer[21]. These DNA sequences are shown in Fig. 2. Note that the plasmid pCPL6: : Tn*7* contains a secondary λ attachment site indicated as PO"B" in the figure (see ref. 2 for a complete description of the construction of pCPL6). —, *E. coli* DNA; ■, Tn*7* DNA; ▨, pBR322 DNA; □, λ DNA.

In two independent experiments, Tn*7* was inserted into the attachment site in pCPL6 by the techniques described in ref. 2; in each case, the site of attachment and the surrounding sequences were identical (see Fig. 2). The following properties of the DNA sequences are of interest.

As with many other transposons, Tn*7* insertions are bounded by short (5 base pair (bp)) directly repeated sequences. The occurrence of a duplication of the pre-existing 5-bp sequence must be a consequence of the mechanism of transposition.

The precise site of insertion occurs in both the *E. coli* chromosome and in Col E1 within a short region of target DNA that contains inverted repeat sequences. Such sequences have alternative pairing configurations and may be represented either by linear duplex DNA or by looped-out cruciform or hairpin structures (see Fig. 3*a*, *b*).

It is difficult to evaluate the importance of these possible secondary structures in Tn*7* transposition. Thermodynamic calculations using the rules of Tinoco[10] suggest that in single-stranded DNA these structures can stably form at 25 °C ($\Delta G = -7.2$ kcal mol^{-1} for structure 3*a* and $\Delta G = -3.6$ kcal mol^{-1} for structure 3*b*). At 37 °C, however, the second of these free energy calculations is unlikely to be a sure guide to the stability of a cruciform structure. There is strong evidence that secondary structures exist in supercoiled duplex DNA (ref. 11), particularly in A+T-rich regions[12]. Bacteriophage λ, which shares many properties with transposable elements, has been shown to require supercoiled DNA for integration[13], and the bacterial chromosome is known to be composed of supercoiled domains[14]. Perhaps site selection by Tn*7* involves structures that directly result from this supercoiling. With other transposable elements, the site of insertion seems not to contain

0028–0836/82/240601—03$01.00

potential cruciforms, but there is a preference for insertion into A + T-rich regions[3].

Tn*7* has unusual asymmetrical termini. Each end contains regions that occur several times in the same orientation and which are closely homologous. They are contiguous in the left end of Tn*7* (boxes 2 and 3 of Fig. 2*a*). These direct repeated sequences of 22 bp were identified by computer analysis[23]; they are more than 75% homologous. In the right end of Tn*7*, they appear in the opposite orientation (boxes 2 and 3, Fig. 2*b*) but are separated by sequences not found at the other end. Figure 3*c* presents a potential hairpin structure that might form by bringing the homologous regions at each end of Tn*7* into register. The asymmetry at the ends of Tn*7* may account for its unique orientation of insertion at the attachment site, as other transposable elements have symmetrical termini and insert in either orientation[3].

Other transposable elements also possess directly repeated sequences within each end. For example, other members of the family of elements that generate 5-bp duplications have the sequence 5'ACGAAAA3' followed by a poorly conserved direct repeat[3]. (A nearly homologous sequence, 5'ACAAAA3', occurs several times within the ends of Tn*7* in box 3.) This common feature may have some role in recognition by the enzymes involved in transposition and, since the duplicated regions (that is, box 3) are more extensive in Tn*7*, this may in part account for the unusually efficient transposition of Tn*7*. Naturally, the nature of the target sequence must also determine the efficiency of transposition.

Tn*7* terminates in the dinucleotide sequence TG . . . CA, as has also been observed in bacteriophage Mu (ref. 3), eukaryotic transposable elements and the provirus forms of retroviruses[15].

A full understanding of the reasons for the extreme specificity of Tn*7* for its attachment site must await further studies. Clearly, the mere presence of a hairpin structure is insufficient, because many chromosomal and plasmid inverted repeat structures (some thermodynamically more stable than the preferred site) are not high efficiency target sequences. Tn*7* may show a preference for insertion near to regions showing DNA sequence homology with its termini, as both insertion sites reported here show ~65% homology with the terminal 12 bp at the ends of Tn*7*; such a preference has also been observed for other transposable elements[3]. Enzymes or mechanisms similar to those involved in bacteriophage integration may be involved, for the regions surrounding the Tn*7* attachment site and a region within one end of Tn*7* show homology to the core sequence of λ *att* (see Fig. 2). Such sequences also occur within the termini of other transposable elements[16], while the element IS*4* favours insertion into a site close to a secondary λ *att* in the *E. coli* chromosome[17].

A better understanding of the factors involved will require a mutational analysis of Tn*7* and its attachment site. Work is under way to determine the DNA sequences of mutations that affect this phenomenon and establish which nucleotides participate in site-specific Tn*7* integration.

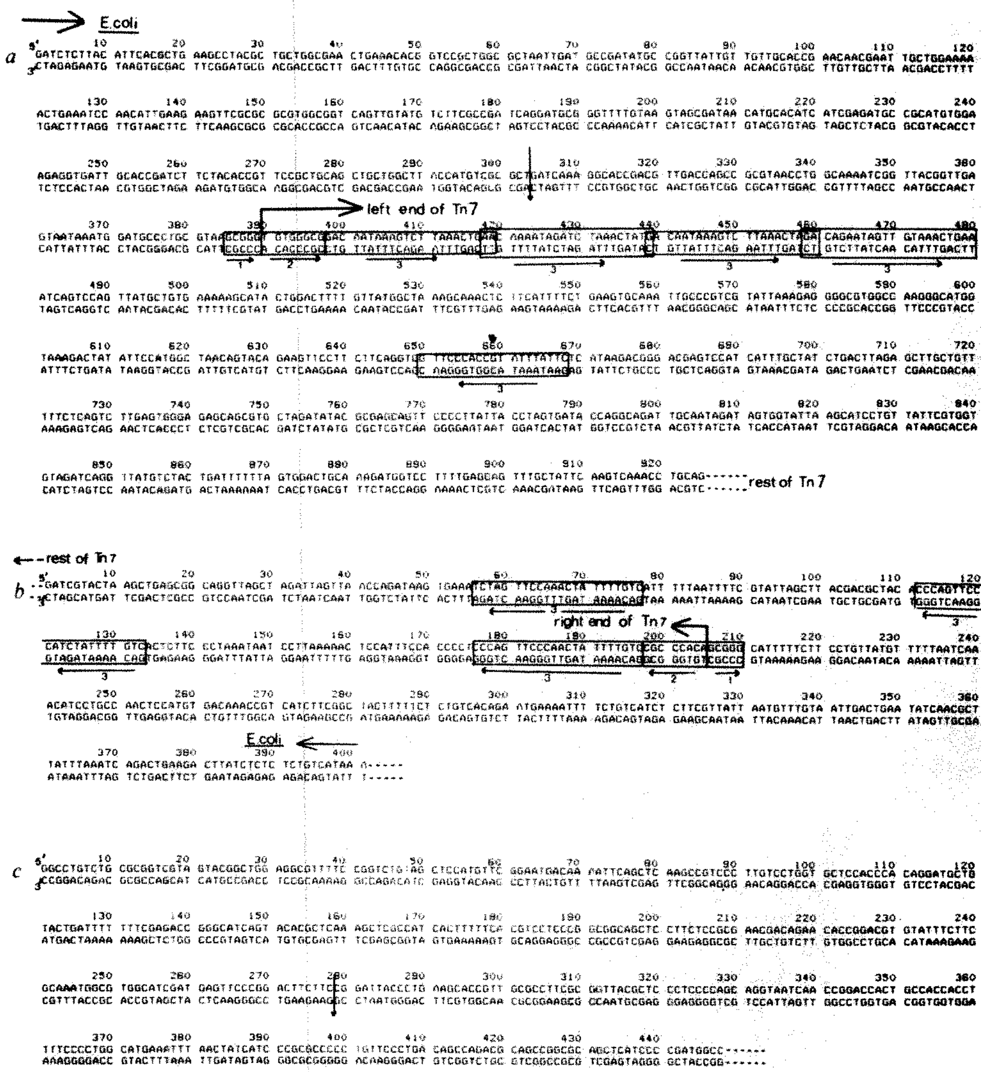


Fig. 2 *a*, *b*, The DNA sequence of the left and right ends of Tn*7* at its attachment site in the plasmid pCPL6::Tn*7*. It has been assumed that the insertion site of Tn*7* in the *E. coli* DNA region of this plasmid is identical to that of the Tn*7* attachment site in the context of the *E. coli* chromosome, an assumption substantiated by restriction enzyme analysis[2]. The left end of Tn*7* terminates at position 390 in *a*. The right end terminates at position 205 in *b*. The DNA sequence of a 13-kb region of Tn*7* DNA between these ends was not determined. Each end is flanked by *E. coli* DNA that surrounds the Tn*7* attachment site (positions 1–389 in *a* and 206–401 in *b*). The *Sau*3A restriction fragment (1 kb) from pCPL6, that includes the Tn*7* attachment site before insertion, begins at nucleotide 304 in *a* (as indicated by the arrow); this fragment was here shown to retain efficient site-specific properties with regard to Tn*7* transposition[2]. A 5-bp sequence of duplicated target DNA (indicated as box 1 in *a* and *b*) flanks the ends of Tn*7* as a direct repeat, as indicated by the arrows. A short G + C-rich region is present at each end of Tn*7* (box 2) in opposite orientation. Other regions of near direct repeat are indicated as box 3 at each end of Tn*7*. An asterisk at position 659 indicates an extra G:C base pair in one of these boxed regions. Certain regions show some DNA sequence homology to the core region (5'GCTTTTTTATACTAA3') of λ *att* (ref. 16) when allowance is made for gaps at some nucleotide positions. These regions are at positions 210–225 (10/15 bp homology) 229–241 (11/15 bp homology) in *b*. A similar sequence occurs at positions 151–138 in the right end of Tn*7* in the opposite orientation (9/15 bp homology). Significant A + T-rich regions that surround the Tn*7* attachment site are at positions 343–369 in *a* (19/27 A + T) and 212–241 in *b* (24/30 A + T). A 12-bp DNA sequence at the Tn*7* attachment site (5'CGTAAGCGGGCA3') shows an 8/12 homology with the terminal 12 bp of Tn*7*. *c* Shows the ColE1 DNA sequence of the *Hae*III G fragment[22], including the Tn*7* insertion site before insertion. This site was deduced by comparison of the right-end junction fragment of ColE1::Tn*7* with the sequence of this region in ColE1 previously obtained by J-I. Tomizawa and H. Ohmori and reproduced here with their kind permission. An arrow between positions 278–279 indicates the Tn*7* insertion site, which is in the same orientation as that shown in the attachment site. The sequence of the right end of Tn*7* in ColE1 was identical with that shown in *b*. Nucleotide positions 262–273 (*c*) show 7/12 bp homology with the terminal 12 bp of Tn*7*.

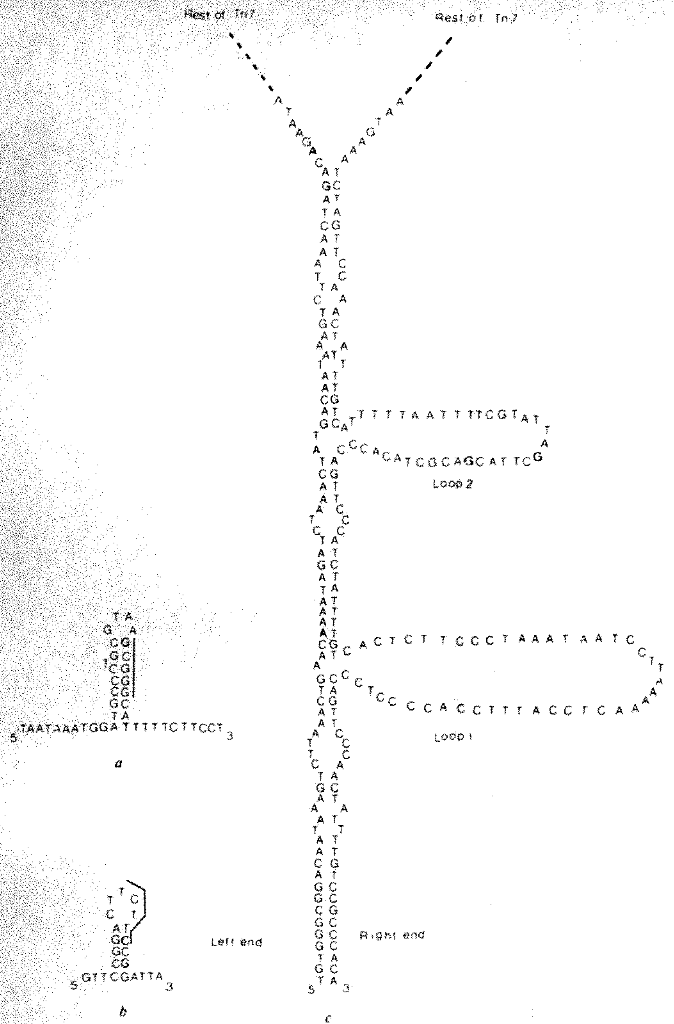


Fig. 3 *a*, The DNA sequence around the Tn*7* attachment site in the plasmid pCPL6 drawn as a potential hairpin structure that might be formed by intra-strand base pairing. This region is flanked by A+T-rich regions. Only one strand is indicated but the other would be present in a mirror image to form a cruciform. The bar marks the pentanucleotide sequence that is duplicated after Tn*7* insertion. *b* Shows a similar, but presumably less stable structure that might form at the Tn*7* insertion site in the plasmid ColE1. *c* Indicates a possible hairpin structure that might form by intra-strand base pairing at the ends of Tn*7*, when boxes 2 and 3 (as indicated in fig. 2*a*, *b*) are brought into register. Loops 1 and 2 are the intervening DNA sequences between the three box 3s in the right end of Tn*7*.

We thank R. P. Anderson and A. Otsuka for comments on the manuscript. The Tn*7* insertion was a gift from M. Saul.

Received 10 July 1981; accepted 28 April 1982.

1. Barth, P. T., Datta, N., Hedges, R. W. & Grinter, N. J. *J. Bact.* **125**, 800–810 (1976).
2. Lichtenstein, C. & Brenner, S. *Molec. gen. Genet.* **183**, 380–387 (1981).
3. Calos, M. P. & Miller, J. H. *Cell* **20**, 579–595 (1980).
4. Bachmann, B. S., Brooks-Low, K. & Taylor, A. L. *Bact. Rev.* **40**, 116–167 (1976).
5. Fennewald, M. A. & Shapiro, J. A. *J. Bact.* **139**, 264–269 (1979).
6. Phillips, G. & Novick, R. P. *Nature* **278**, 476–478 (1979).
7. Bolivar, F. *et al. Gene* **2**, 95–113 (1977).
8. Miki, T., Hiraga, S., Nagata, T. & Yura, T. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5099–5103 (1978).
9. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
10. Tinoco, I., Uhlenbeck, O. C. & Levine, M. D. *Nature* **230**, 362–367 (1971).
11. Lilley, D. J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6468–6472 (1980).
12. Panayototas, N. & Wells, R. D. *Nature* **289**, 466–470 (1981).
13. Gellert, M., Mizuuchi, K., O'Dea, M. & Nash, A. *Proc. natn. Acad. Sci. U.S.A.* **73**, 3872–3876 (1976).
14. Worcel, A. & Burgi, E. *J. molec. Biol.* **71**, 127–147 (1972).
15. Temin, H. *Cell* **21**, 599–600 (1980).
16. Zieg, J. & Simon, M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4196–4200 (1980).
17. Hubermann, P., Klaer, R., Kuhn, S. & Starlinger, P. *Molec. gen. Genet.* **175**, 1127–1134 (1979).
18. Rothstein, R. J., Lan, L. F., Bahl, C. P., Narang, S. A. & Wu, R. *Meth. Enzym.* **68**, 98–109 (1979).
19. Benton, W. D. & Davis, R. W. *Science* **196**, 180–182 (1977).
20. Bentley, D. L. & Rabbitts, T. H. *Nature* **288**, 730–733 (1980).
21. Schreier, P. H. & Cortese, R. J. *J. molec. Biol.* **129**, 169–172 (1979).
22. Tomizawa, J-I., Ohmori, H. & Bird, R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 1865–1869 (1977).
23. Staden, R. *Nucleic Acids Res.* **5**, 1013–1015 (1978).

Structural analysis of the prolactin gene suggests a separate origin for its 5′ end

Nancy E. Cooke* & John D. Baxter

Howard Hughes Medical Institute Laboratory,
Department of Medicine, Biochemistry and Biophysics; and
Metabolic Research Unit, University of California, San Francisco,
California 94143, USA

Prolactin, growth hormone (GH) and placental lactogen constitute a family of polypeptide hormones which share certain structural[1–7] and functional[8] features. The prolactin and GH genes evolved from a gene duplication about 400 million years ago[9] and, in humans, segregated onto chromosomes 6 and 17 respectively[10,11]. We expand here previous reports[12,13] on the general organization of the rat prolactin gene, *rPRL*, assign its haploid gene number and extend its DNA sequence analysis. We postulate that a direct repeat flanking exon I may be the remnant of an insertional event.

A single plaque hybridizing to a *rPRL* cDNA[9] probe was isolated after screening 500,000 plaques of a λCharon4A library of Hooded rat chromosomal DNA. Digestion of the isolated recombinant phage DNA with *Pst*I or *Hin*dIII endonuclease resulted, in each case, in two fragments that hybridized to the *rPRL* cDNA probe (Fig. 1, lanes 3, 5). The four fragments were individually subcloned into pBR322 and used in conjunction with the recombinant phage to map and sequence *rPRL*. The single recombinant phage and its subclones encoded the

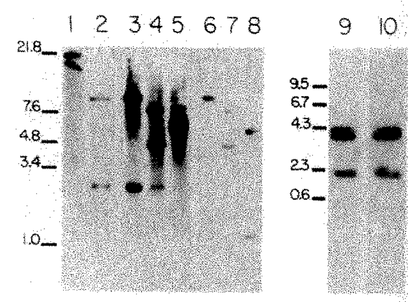


Fig. 1 Detetion of *rPRL* and distribution of coding and repetitive sequences among its restriction fragments. Lanes 1, 2: Sprague–Dawley rat liver DNA digested with *Bam*HI and *Pst*I respectively, electrophoresed on a 0.8% agarose gel, transferred to nitrocellulose and hybridized to a ^{32}P-labelled *rPRL* cDNA probe[9]. Lanes 3–5: DNA fragments from the Hooded *rPRL* recombinant λCharon4A clone digested with *Pst*I, *Pvu*II and *Hin*dIII respectively, separated on a 0.8% agarose gel and hybridized to the ^{32}P-labelled *rPRL* cDNA probe. Lanes 6–8: as for lanes 3–5, but hybridized to ^{32}P-labelled total rat DNA after melting off the previous probe. Repetitive and coding sequences are both present in the larger of the two *Pst*I fragments and in the 5.6 kbp *Hin*dIII fragment, but the 1.5 kbp *Hin*dIII fragment contains only repetitive DNA and is located in the 3′-flanking region of the gene. Lanes 9, 10: autoradiograms of an identical blot of *Eco*RI-digested *rPRL* subclone pRPRL–*Pst*Q (see Fig. 2*c*) hybridized to ^{32}P-labelled rat DNA and then to a probe made from the 3′-repetitive region respectively. The positions of molecular size markers are indicated. Templates for probes were labelled internally with ^{32}P-dNTPs to a specific activity of 1×10^8 c.p.m. μg^{-1} using random primers[16] or nick translation[29]. Total DNA probes used to map repetitive sequences were synthesized by internally labelling rat DNA sheared to 500 bp. These probes have been shown to detect sequences in >50–100 copies per haploid genome in the hybridization conditions used[30]. *rPRL* was isolated from a Hoodel rat λCharon4A library (kindly provided by G. Paige) composed of partial *Eco*RI-digested rat DNA fragments ligated into phage arms. Established techniques were used for clone identification[31] and amplification[32]. Restriction endonucleases were purchased from New England Biolabs, Boeringer-Mannheim and Bethesda Research Laboratories, and used as recommended.

* Present address: Endocrine Section, Department of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.

0028–0836/82/240603—04$01.00

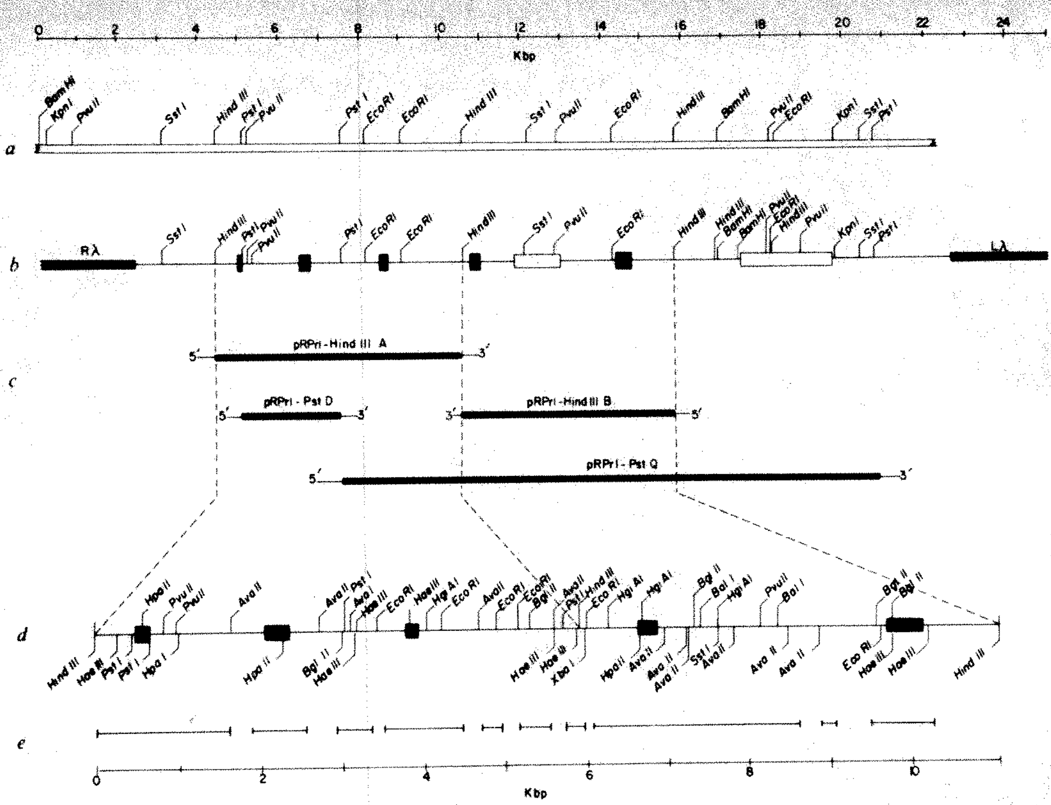


Fig. 2 Organization of *rPRL*. Restriction maps of *a*, the prolactin gene in Sprague–Dawley DNA; and *b*, the λCharon4A-*rPRL* recombinant clone. R and L refer to the right and left phage arms respectively, exons are indicated by solid rectangles and repetitive regions that hybridized to total DNA probes by white rectangles. *c*, Location and orientation in pBR322 of the λ-*rPRL* subclones used for detailed mapping and sequencing. *d*, Detailed structural map of the area sequenced and *e*, the extent of sequencing, indicated by the bracketed lines. *a–c* are drawn to the kbp scale on the top, and *d* and *e* to the scale on the bottom. The subclones indicated in *c* were constructed by ligating *Pst*I or *Hin*dIII digests of the λCharon4A-*PRL* recombinant into 5′-dephosphorylated (bacterial alkaline phosphatase, P-L Biochemicals) *Pst*I-or *Hin*dIII-digested pBR322 using T4 DNA ligase (New England Biolabs). *Escherichia coli* HB 101 were made competent by $CaCl_2$ treatment[33], transformed with the ligation mixtures, and recombinant colonies containing *rPRL* restriction fragments identified by hybridization to ^{32}P-labelled *rPRL* cDNA[34].

entire *rPRL* gene (Fig. 2*b*). The location, size and orientation of all the pBR322 subclones are shown in Fig. 2*c*.

The map of the isolated Hooded rat prolactin gene was identical to the *rPRL* mapped in the Sprague–Dawley genome (Fig. 2*a*, *b*). Both maps were constructed by hybridizing ^{32}P-labelled restriction fragments from specific regions of *rPRL* cDNA to Southern blots[14] containing restriction fragments of the *rPRL*-λCharon4A recombinant clone, or digests of

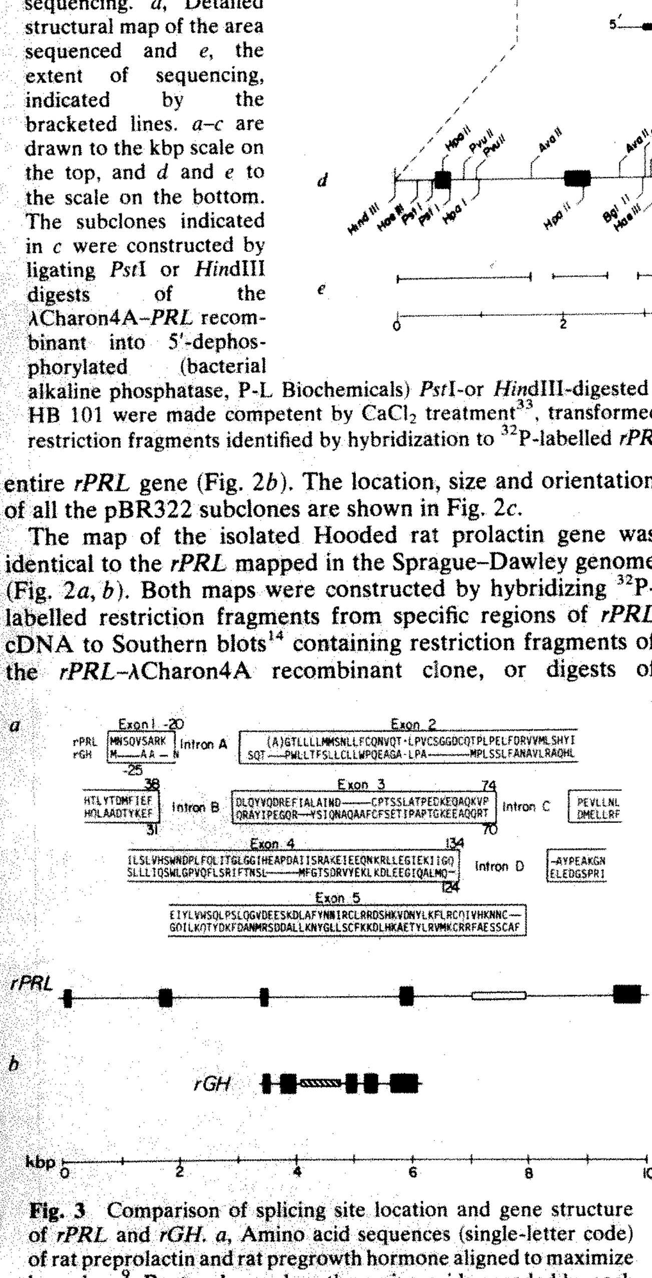


Fig. 3 Comparison of splicing site location and gene structure of *rPRL* and *rGH*. *a*, Amino acid sequences (single-letter code) of rat preprolactin and rat pregrowth hormone aligned to maximize homology[9]. Rectangles enclose the amino acids encoded by each exon. The amino acids located at the carboxy terminus of the coding region of each exon are numbered. In *rPRL*, the codons for alanine (A) or glycine (G) indicated at the beginning of exon 2 are actually split by intron A (see text for discussion). *b*, Comparison of the sizes of *rPRL* and *rGH*. Exons are indicated by solid rectangles, dispersed repetitive sequences by a white rectangle in *RPRL* and a hatched rectangle in *rGH*. Both genes are shown to the scale indicated at the bottom of the figure.

Sprague–Dawley liver DNA. The 10.5 kilobase pair (kbp) *rPRL* was found to consist of five exons and four introns (Fig. 2*b*, *d*), similar to the pattern found in the rat GH gene (*rGH*)[15,16], and in good agreement with the size estimates predicted by heteroduplex mapping of the Sprague–Dawley *rPRL*[17]. The sizes of the *Pst*I genomic *rPRL* fragments and those from the *rPRL* recombinant phage DNA were identical (Fig. 1, lanes 2, 3). There were no *Bam*HI sites in the cloned gene, and only one hybridizing *Bam*HI genomic fragment was seen (Fig. 1, lane 1). Previously only one *rGH* was mapped in the rat genome[15,16], although multiple GH genes are present in the human genome[18]. It seems reasonable to conclude that there is only one *rPRL*, although the presence of additional prolactin genes embedded in an identical surrounding DNA environment is possible. A previously mapped and partially sequenced Sprague–Dawley *rPRL* lacks 2.55 kbp of DNA in intron D[12,13]. We found no evidence for this structure and conclude that it may represent a polymorphic variant, cloning artifact, or mapping error.

DNA sequence analysis has shown that the slicing junctions f introns A, B, C and D of *rPRL* are of the same class[19] and occur at analogous positions to those determined for *rGH* when both coding structures are aligned to maximize homology (Fig. 3*a*). Conservation of intron location in coding regions is frequently seen among individual members of gene families[19,20]. The conservation of splicing sites was particularly striking between *rPRL* and *rGH* as they retain only about 25% amino acid homology and *rPRL* is five times larger than *rGH* (Fig. 3*b*). This conservation of splicing site locations substantiates the hypothesis that *rPRL* and *rGH* evolved from a common precursor.

Although the similar overall structures of *rPRL* and *rGH* implied a common precursor, lack of homology and dissimilar lengths between exons I of these genes suggested that the 5′ ends may have had independent origins. A direct decanucleotide repeat which flanks exon I of *rPRL*, occurs at bases −77 to −68 in the 5′-flanking region and again in intron A at bases 340–349 (boxed in Figs 4, 5*a*). Similarly, an 80% homologous

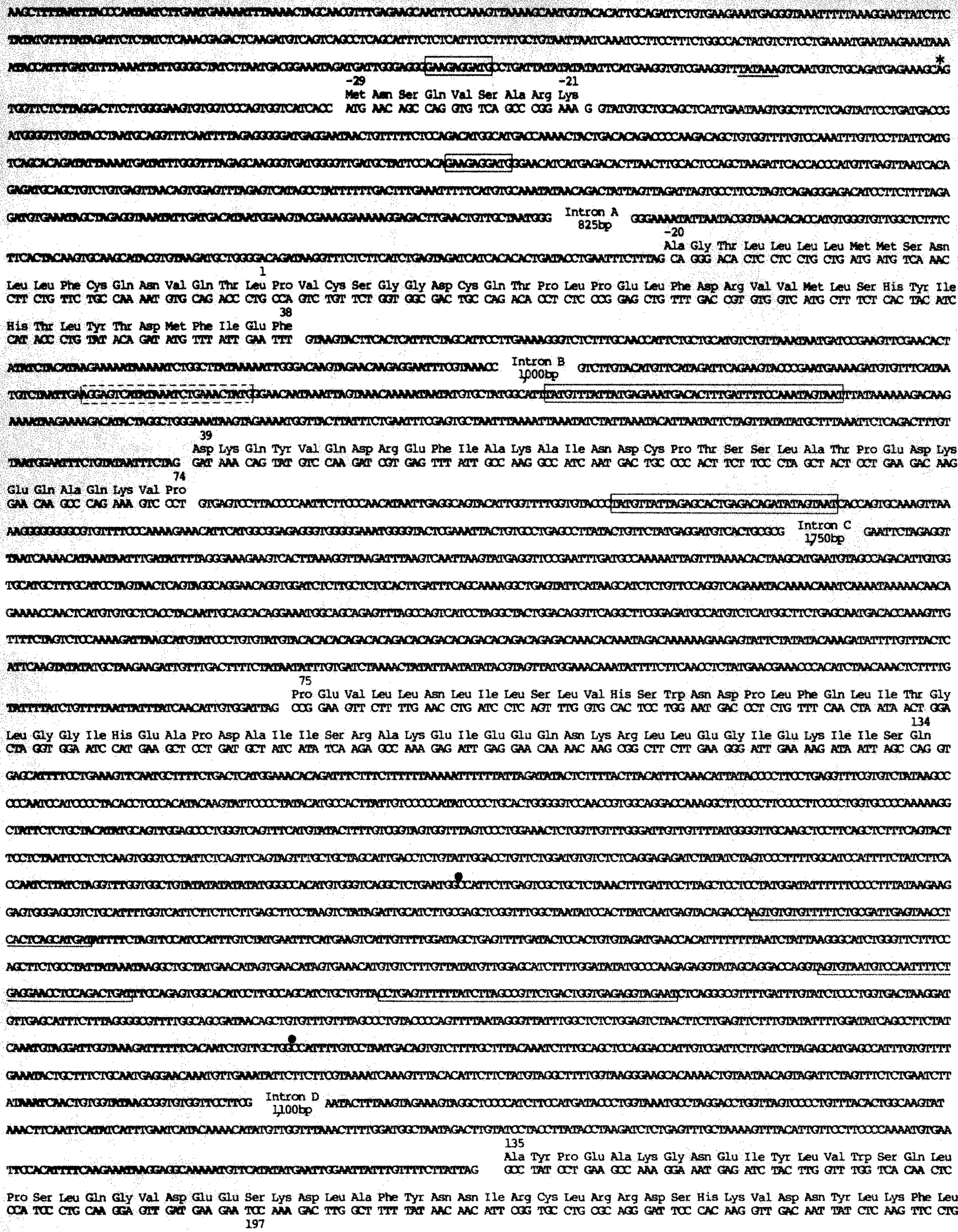


Fig. 4 DNA sequence of *rPRL*. The sequence begins 423 bp 5′ to the cap site, includes 5,070 bp of the gene and ends near the poly(A)addition site. The numbered amino acids of rat preprolactin are shown. The TATAAA sequence is underlined once, the putative capping site is starred. Two sets of direct repeats are enclosed in rectangles. A dashed rectangle encloses an area of homology with intron B of *rGH*. The dots in intron D demarcate the approximate limits of hybridization to total rat DNA probes and three areas of homology to human Alu family consensus sequences are underlined. The lengths of areas not presented are indicated. DNA sequencing was carried out by the chemical cleavage technique of Maxam and Gilbert[35], as previously modified[9]. Approximately 90% of the sequences shown were either determined twice or on both strands. DNA sequences were analysed in a PDP/1160 computer (Digital Equipment Corp.) using programs provided by H. Martinez[36].

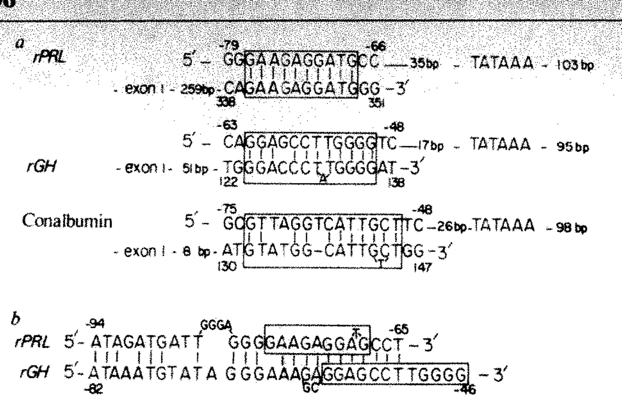


Fig. 5 Direct repeats flanking exon I of *rPRL*, *rGH* and conalbumin gene. *a*, Diagram of an isolated segment of the 5′-flanking region, promoter site, first exon and first intron of three genes. Rectangles enclose a decanucleotide repeat (100% homology) in *rPRL*, a 12 bp repeat (one mismatch, one insertion/deletion) in *rGH*[16] and a 14 bp repeat (79% homology, two insertions/deletions) in conalbumin gene[23]. The number of base pairs separating the 5′ repeat and TATAAA, TATAAA and the 5′ border of exon I, and the 3′ border of exon I and the 3′ repeat are indicated. The actual base number of the repeat units (cap site is +1) are indicated above or below the appropriate nucleotides. *b*, Shows an area of homology in the 5′-flanking regions of *rPRL* and *rGH*. The 5′ direct repeat in each gene, shown in *a*, is enclosed by a rectangle in *b*.

12 basepair (bp) repeat, different in sequence from the *rPRL* repeat, flanks exon I of *rGH* (Fig. 5*a*), occurring at bases −61 to −50 and 124 to 136[14]. In both genes, the repeats encompass TATAAA (a probable promoter region), the mRNA cap sites and exon I. The presence of these direct repeats may indicate that the encompassed sequence was inserted into these genes by mechanisms proposed for prokaryotic, viral and eukaryotic insertion sequences and transposons[21]. The insertion of mobile genetic elements characteristically results in a duplication of host target site DNA bracketing the inserted DNA. Although the sequences of *rPRL* and *rGH* are divergent, the DNA sequences encompassing the more 5′ of the direct repeats flanking exon I of r*PRL* and *rGH* were highly homologous (76%) (Fig. 5*b*). In both genes this homologous region was located in the approximate area that, in many other eukaryotic genes[22], contains the CAAT homology unit, which is lacking in both *rPRL* and *rGH*. Sixty-nine bp upstream from the 5′ copy of the decanucleotide direct repeat in *rPRL* was a second and presumably non-functional promoter sequence (AATAAA) preceded 42 bp upstream by a CAAAT sequence. *rGH* also has a second AATAAA, 21 bp 5′ to its repeat, but lacks the upstream CAAT sequence. These data suggest that the proposed insertion of DNA encoding exon I and its 5′-flanking region occurred just 3′ to the original promoter sequences of these genes in a homologous 'hotspot'[21] for integration. The proposed insertion of exon I may not be unique to *rGH* and *rPRL*. For example, a direct repeat flanking exon I in the chick conalbumin gene[23] was detected (Fig. 5*a*). As in *rPRL* and *rGH*, the direct repeats encompass the TATAAA sequence and the entire coding unit of the first exon. As exons I of *rPRL* and *rGH* are non-homologous and of different lengths, we concluded that they may have been inserted independently into already distinct prolactin and GH genes. Such an event could place new regulatory signals 5′ to these genes and facilitate divergence of regulation and function. The 73% nucleotide homology between the signal peptides of human and rat prolactin[7] suggested that the insertion of exon I, which partially encodes this region of the protein, occurred before the divergence of rat and human. Conservation of the decanucleotide repeats during further evolution may have been a random event or may indicate some functional constraint on their structure.

Introns A, B, C and D interrupt the coding regions of *rPRL* between codons −20/−19, 38/39, 74/75 and 134/135 respectively (Figs 3, 4). The precise positions of the splicing sites were assigned by locating the characteristic G–T at the 5′ end of the intron and A–G at the 3′ end of the intron[24], as well as by optimizing homology to U1 RNA[25–27]. The actual position of the 3′-splicing junction of intron A is ambiguous[13]. There are two possible splicing sites, one that includes an alanine (GCA) in the signal peptide (Fig. 4), and another that excludes this alanine and splices directly to glycine (GGG) at amino acid position −19. The former preserves the greatest complementarity with rat U1 RNA (60% compared with 40%). Two separate *rPRL* cDNAs sequenced from oestrogen-stimulated rats contained the alanine codon[12,28], while a third *rPRL* cDNA clone isolated from rat pituitaries maximally stimulated by oestrogen injections plus hypothalmic ablation did not contain this codon[9]. The identification of both mRNA species is consistent with the prediction of alternative splicing of the single *rPRL* gene. No similar situation occurs in the *rGH*. Whether the anomalous nature of this splice site relates to a separate evolutionary history of exon I in *rPRL* and *rGH* remains speculative.

Two regions of dispersed repetitive sequences were present in the cloned 20 kbp genomic fragment containing *rPRL*. Their presence and location were determined by Southern blot hybridization using ^{32}P-labelled total rat DNA as probe (Fig. 1, lanes 6–9). Results of more detailed mapping localized these two repetitive DNA sequences to intron D and to the 3′-flanking region (Fig. 2*b*). These two repetitive regions do not hybridize to each other, or to the repetitive DNA in intron B of *rGH*. There are therefore at least three discrete families of dispersed repetitive elements in the rat genome represented in and around these two related genes. Insertion of these repetitive sequences may serve rapidly to expand intron DNA and may limit interaction between the two genes, thus facilitating their structural divergence.

We thank Ilana Mittman for technical assistance, and Drs Stephen Liebhaber, John Shine and Y. W. Kan for advice, assistance and support.

Received 19 October 1981; accepted 14 April 1982.

1. Catt, K. J., Moffat, B. & Niall, H. D. *Science* **157**, 321 (1976).
2. Li, C. H., Dixon, J. S., Lo, T. B., Pankov, Y. M. & Schmidt, K. D. *Nature* **224**, 695 (1969).
3. Sherwood, L. M. *Proc. natn. Acad. Sci. U.S.A.* **58**, 2307–2314 (1967).
4. Niall, H. D., Hogan, M. L., Sauer, R., Rosenblum, I. Y. & Greenwood, F. C. *Proc. natn. Acad. Sci. U.S.A.* **68**, 866–869 (1971).
5. Seeburg, P. H., Shine, J., Martial, J. A., Baxter, J. D. & Goodman, H. M. *Nature* **270**, 486–494 (1977).
6. Shine, J., Seeburg, P. H., Martial, J. A., Baxter, J. D. & Goodman, H. M. *Nature* **270**, 494–499 (1977).
7. Cooke, N. E., Coit, D., Shine, J., Baxter, J. D. & Martial, J. A. *J. biol. Chem.* **256**, 4007–4016 (1981).
8. Niall, H. D. *et al. Recent Prog. Horm. Res.* **29**, 387–416 (1973).
9. Cooke, N. E., Coit, D., Weiner, R. I., Baxter, J. D. & Martial, J. A. *J. biol. Chem.* **255**, 6502–6510 (1980).
10. Owerbach, D., Rutter, W. J., Cooke, N. E., Martial, J. A. & Shows, T. B. *Science* **212**, 815–816 (1981).
11. Owerbach, D., Rutter, W. J., Martial, J. A., Baxter, J. D. & Shows, T. B. *Science* **209**, 289–291 (1980).
12. Gubbin, E. J., Maurer, R. A., Legrimini, M., Erwin, C. R. & Donelson, J. E. *J. biol. Chem.* **255**, 8655–8662 (1980).
13. Maurer, R. A., Erwin, C. R. & Donelson, J. E. *J. biol. Chem.* **256**, 10524–10528 (1981).
14. Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).
15. Page, G. S., Smith, S. & Goodman, H. M. *Nucleic Acids Res.* **9**, 2087–2104 (1981).
16. Barta, A., Richards, R., Baxter, J. D. & Shine, J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4867–4871 (1981).
17. Chien, Y. & Thompson, E. B. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4583–4587 (1980).
18. Goodman, H. M. *et al.* in *Mobilization and Reassembly of Genetic Information* (eds Scott, W. A., Werner, R., Joseph, D. R. & Schultz, J.) 155–179 (Academic, New York, 1980).
19. Sharp, P. A. *Cell* **23**, 643–646 (1981).
20. Efstratiadis, A. *et al. Cell* **21**, 653–668 (1980).
21. Calos, M. P. & Miller, J. H. *Cell* **20**, 579–595 (1980).
22. Benoist, C., O'Hare, K., Breathnach, R. & Chambon, P. *Nucleic Acids Res.* **8**, 127–142 (1980).
23. Cochet, M. *et al. Nature* **282**, 567–574 (1979).
24. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. & Chambon, P. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4853–4857 (1978).
25. Seif, I., Khoury, G. & Shar, R. *Nucleic Acids Res.* **6**, 3387–3398 (1979).
26. Rogers, J. & Wall, R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1877–1879 (1980).
27. Lerner, M. Boyle, J., Mount, S., Wolin, S. & Steitz, J. *Nature* **283**, 220–224 (1980).
28. Taylor, W. L., Collier, K. J., Weith, H. L. & Dixon, J. E. *Biochem. biophys. Res. Commun.* **102**, 1071–1077 (1981).
29. Maniatis, T., Jeffrey, A. & Kleid, D. G. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1184–1188 (1975).
30. Shen, C. J. & Maniatis, T. *Cell* **19**, 379–391 (1980).
31. Benton, W. D. & Davis, R. W. *Science* **196**, 180–182 (1977).
32. Maniatis, T. *et al. Cell* **15**, 687–701 (1978).
33. Dagert, M. & Ehrlick, D. S. *Gene* **6**, 23–28 (1979).
35. Grunstein, M. & Hogness, D. S. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3961–3965 (1975).
35. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564 (1977).
36. Martinez, H. M., Bock, S. A. & Katzung, B. *DNA Programs Manual* (Department of Biochemistry and Biophysics, University of California, San Francisco, 1980).

BOOK REVIEWS

The bones of palaeoanthropology

Andrew Hill

SOMETIMES in science a totally new class of things becomes a legitimate object of study, not through having previously been undiscovered, but through previously mundane items acquiring relevance. Recently, palaeontologists and archaeologists formed an interest in natural accumulations of modern broken bones as an aid to understanding the origins and significance of vertebrate fossil assemblages. The question of what happens to animals between death and possible fossilization has even been blessed with the status of its own name — taphonomy — although it is little more than a necessary prior consideration of any reliable palaeoecological work. A number of fairly discrete centres of interest have coalesced to constitute the activity as it is developing today, and one of the most important of these relates to Plio-Pleistocene hominids in sub-Saharan Africa. Amongst scientists involved, C.K. Brain stands out as the pioneer; this impressive book is a statement of his investigations.

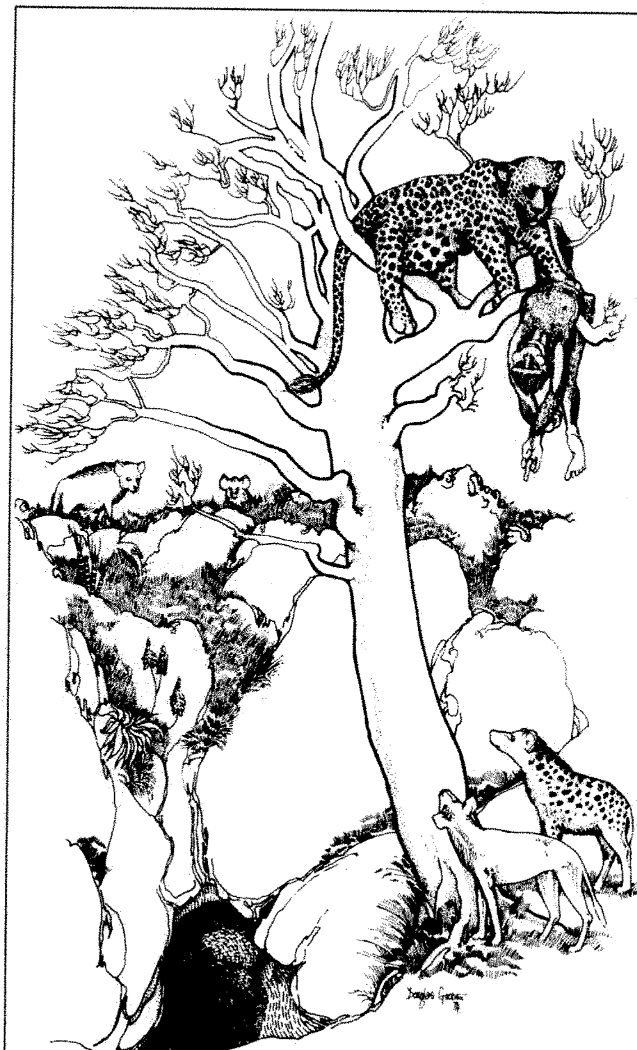
Before the discovery of many early hominids in eastern Africa most of the evidence for human evolution came from a few caves in the far south. The hominids and associated bone accumulations became controversial, mainly because Raymond Dart plausibly suggested most of the bones were tools and weapons fashioned by australopithecines. He also claimed evidence of violence, offered not only to other species but to other hominids. This is the aggressive view propagated by Robert Ardrey. But such ignoble savagery is no longer in vogue and nature's teeth and claws are less red than they were. Beware of the dogma; could it instead have been cave carnivores that produced the bone remains in feeding, or was it really the hominids that were responsible for — as Philip Tobias put it — the "lion's share"? This is an important matter relevant to our sense of the behaviour of early man, and maybe to our feelings about our own nature today. It is around this question that *The Hunters or the Hunted?* is structured.

The Hunters or the Hunted? An Introduction to African Cave Taphonomy. By C.K. Brain. Pp.365. ISBN 0-226-07089-1. (University of Chicago Press: 1981.) £24.50, $45.50.

Brain attempts to provide an answer by looking in some detail at a large spectrum of contemporary creatures that accumulate bones, both in caves and in more open situations. Consequently the first part of the book is devoted to accounts of the extensive observations and the results of experiments on the feeding and food remains of modern human beings and their dogs, and of cave carnivores such as leopards and hyaenas. Porcupines are curious collectors of bones which are also dealt with, as are birds such as eagles and owls which contribute quantities of small vertebrate material to cave sediments. In addition there is a discussion of the middens of later prehistoric human cave dwellers in southern Africa. A number of aspects of the accumulations are investigated, such as the various kinds of bone breakage produced by these different factors, the size and numbers of the prey, and the proportional representation of different skeletal parts in the collections. Based upon such features as these Brain successfully establishes criteria for distinguishing between most of the collecting agencies.

Next follows an incidental but most useful summary of information about all fossil animal species from the Sterkfontein Valley caves. The last section takes each of these cave systems in turn — Sterkfontein, Swartkrans and Kromdraai — and within them each stratigraphic member, in order to assess the origins of the bone accumulations. In this the comparative information presented in Part One is employed. An extremely valuable element is a complete taxonomic listing of all the fossil vertebrate specimens from the deposits, and there is an historical description of the discovery and subsequent work at the sites. Notes are also provided on the other South African australopithecine caves — Taungs and Makapansgat — that Brain has not studied in such detail. All of this is thoroughly documented, supported by an appendix of 121 tables, and there are many illustrations including some unfortunately rather murky photographs.

New branches of activity in natural science have a tendency to reach their anecdotage before maturity. Is this a necessary stage of development, or an avoidable fault? It is certainly a feature of the first part of the book, where although many of the accounts add greatly to the interest, they also seem to overload the argument with trivial facts that add nothing substantial or lead nowhere. The data — like the



A stylized impression of the Swartkrans leopard hypothesis, a theory which might account for the abundance of hominid remains in the Swartkrans cave. An australopithecine is being consumed in a *Celtis* tree overhanging a cave, while hyaenas of several species wait hopefully below. (Picture courtesy of C.K. Brain.)

food remains that Brain describes — are not fully digested, but they will provide a rich resource for the depredations of other scientific scavengers. It would be useful for some of these to be of a more statistical inclination. There are, admittedly, methodological difficulties with the analysis of the kind of numerical information that Brain has collected; conventional statistical manipulations are always possible, but it is not always clear that the answers mean much in palaeoecological terms. Certainly his tentative approach, drawing essentially qualitative and specific conclusions through closely argued reasoning, yet presenting the original numerical data in full detail, is better than making more suspicious assertions based on operations only superficially possessing greater mathematical rigour. However, a statistical attack on the problem might be beneficial, for I feel that the data contain more information than the author has at present extracted from them; not that his conclusions are slight.

Brain's meticulous reappraisal seems conclusively to show little evidence of early hominids flailing at each other and at other animals with bone clubs or other weapons. Most of the assemblages are postulated to be the product of hyaenas and large felids such as leopards and the extinct *Dinofelis* that controlled the caves, probably a result of opportunistic predation at hominid sleeping sites. But between Sterkfontein Members 4 and 5 came a change, corresponding to that crucial step in human evolution when man put the cat out for the first time. The later bone collection has an exclusively human character, showing that we had at last attained some dominance over other predators. Brain's work, however, supports the idea that initially the hominids relied heavily upon scavenging before increasing intelligence and the beginning of technology permitted the greater development of hunting abilities.

It is gratifying to see these very early hominids regarded as animals on their own terms, as species which are extinct and, no doubt, ones which had their own appropriate mode of life and behaviours. Identifying them too simply with models founded either on modern technologically undeveloped peoples or on modern pongids sometimes masks this fact, and it can be avoided by dealing with the matter from a more fundamental ecological basis, as Brain does here.

The Hunters or the Hunted? is a very important book for palaeoanthropology. It presents the first thorough analysis of the Sterkfontein Valley assemblages, contributes significantly to the resolution of lingering controversies and, by placing the old information in a fresh perspective, enables new and more sophisticated questions to be asked not only of the South African material but of similar assemblages elsewhere. Another contribution is that it reinforces the recent change in feelings as to what constitutes data, for the value of looking at fossil and contemporary bones as closely as this is clear. Brain urges the necessity of recovering fossils with a high regard for subtle detail. I hope that excavators of any vertebrate fossil site will be persuaded to follow his advice and pay more attention to these features of bone accumulations that have been previously neglected; for taphonomy can be a powerful tool in elucidating the problems of fossil assemblages, especially when handled with the care and caution that Brain brings to the subject. □

Andrew Hill is a Research Fellow at Harvard University.

Journals' review issue 1982

On October 7th *Nature* will publish a second review supplement devoted to science journals. Last year's supplement, covering journals first published between January 1978 and May 1980, appeared in *Nature* **293**, 341-369; see p.343 of that issue for details.

Criteria for inclusion of a journal in the 1982 issue are that:

- the first number appeared, or the journal was re-titled, between June 1980 and May 1981;
- it is published at least three times a year;
- the main language used is English.

Broadly, periodicals of professional interest to scientists will be considered for review, with the exception of abstracts' journals.

In addition it is hoped to cover publicly available newsletters, first published between January 1978 and May 1981, in that issue.

Publishers and societies are invited to submit four sample issues of periodicals satisfying the above criteria, *including the first and most recent numbers*, to the Review Editor, *Nature*, 4 Little Essex St, London WC2R 3LF, England (London 836 6633 ext 2570) as soon as possible.

Structure in (and of) the neurosciences

Margaret Boden

Meaning and Purpose in the Intact Brain: A Philosophical, Psychological, and Biological Account of Conscious Processes. By Robert Miller. Pp.239. ISBN 0-19-857579-3. (Clarendon/Oxford University Press: 1982.) £20, $34.50.

ROBERT Miller is a neuroscientist dissatisfied with neuroscience. It has not produced a general theory of consciousness, nor does it regard questions about consciousness, purpose, meaning, understanding or cognition as central. It has contributed little to our understanding of the behaviour of the whole organism, the overall patterns and strategies of living. Instead of "it" one should perhaps put "they", for there are a bewilderingly diverse and narrowly specialized set of neurosciences. In general, neuroscientists ignore psychology and deliberately eschew philosophical speculation.

One cannot but agree with Miller's criticisms of neuroscience. One may even applaud his aim of synthesizing brain research, psychology and philosophy in the search for a theory of consciousness. Nevertheless, despite its laudable intentions, this book is a failure. It does not help us progress significantly towards a theoretical understanding of meaning, purpose, cognition or consciousness.

This failure is grounded in Miller's reliance on an outmoded method of argument, and — ironically, given his critique of professional specialism — in his ignoring recent concepts and forms of argument which offer much greater promise of bringing neuroscience closer to psychology (and philosophy).

Characteristically, Miller argues from structure to function or vice versa. So he gives an account of consciousness in terms of "omniconnected networks", sets of neurones showing synaptic plasticity and influenced by transmitter substances of various kinds. Doubtless such networks are involved. But, as computer science has helped to show, the notion that specific cognitive functions can be attributed to specific cerebral mechanisms on structural grounds alone is mistaken.

Moreover, computational models of psychological processes provide concepts better suited to express specific cognitive functions than the vague psychological terminology used by Miller. Marr's physiologically-informed work on the visual system, for instance, is nowhere mentioned (though his earlier work on the neocortex is cited). Marr argued that neuroscientists will need to specify the computational tasks that the brain must (or might) be performing, before discovering which physiological mechanisms are performing them. Whether or not one regards this claim as overly sweeping, it is effecting radical changes in our view of physiological psychology. Miller's book is gravely flawed by its failure to consider the usefulness of computational ideas in neuroscience. □

Margaret Boden is Professor of Philosophy and Psychology at the University of Sussex. Her books include Purposive Explanation in Psychology *(Harvard University Press, 1972) and* Artificial Intelligence and Natural Man *(Basic Books, 1977).*

The ore deposit—tectonics connection

Frederick J. Sawkins

Mineral Deposits and Global Tectonic Settings. By A.H.G. Mitchell and M.S. Garson. Pp.404. ISBN 0-12-499050-9. (Academic: 1982.) £23.60, $48.50.

WITH *Mineral Deposits and Global Tectonic Settings,* Mitchell and Garson are to be congratulated on providing us with a most impressive piece of scholarship. The interrelationships between tectonics and mineral resources have attracted increasing attention from earth scientists during the past decade, and their book represents by far the most competent and complete analysis of the subject which has emerged so far.

With commendable thoroughness the authors deal successively with hotspot and rifting environments, passive margins and continental margins, oceanic settings, subduction environments and collision settings, and finally transform faults, both in terms of their geology and their associated mineral deposits. The subject is approached in a no-frills, workmanlike style and few stones are left unturned, especially with respect to discussion of the various geological sub-environments within the major types of plate boundary settings. The 164 illustrations that accompany the text are, in the main, well chosen and illuminating, and the bibliography is not only extensive, but consists mainly of papers published within the past decade. The treatment is truly world-wide in scope, although the number of examples of geological terrains and mineral deposits taken from Communist-bloc countries is understandably rather small.

In view of the foregoing, it is perhaps less than charitable to focus on what I regard as the relatively minor shortcomings of the volume. The great challenge inherent in such an undertaking is the achievement of balance, presumably guided, at least in part, by economic significance. Here the authors' greater experience in Africa and the Far East, as opposed to the Cordillera of North and South America, is discernible. For example, considerable emphasis is placed on alkaline complexes and ophiolites, and important Cordilleran deposits such as polymetallic skarns and epithermal precious metal deposits are given short shrift. Porphyry copper deposits are also dealt with in surprisingly few pages considering their pre-eminent economic importance and well-defined tectonic setting. Also, although tonnage and grade figures are given for some major deposits, the distinction between mere mineral occurrences, small deposits and major orebodies is not at all clear in many instances.

This book, however, represents mandatory reading for anyone interested in mineral deposits–tectonics relationships. It will be a valuable resource to those concerned with any aspect of earth resources, tectonics or, for that matter, the regional geology of our planet. □

Frederick J. Sawkins is a Professor in the Department of Geology and Geophysics at the University of Minnesota.

Plants in the sea

G.E. Fogg

Marine Botany. By Clinton J. Dawes. Pp.628. ISBN 0-471-07844-1. (Wiley: 1981.) £33.50, $59.85. *The Biology of Seaweeds.* Edited by Christopher S. Lobban and Michael J. Wynne. Pp.786. UK ISBN 0-632-00672-2; US ISBN 0-520-04585-8. (Blackwell Scientific/University of California Press: 1982.) £45, $85.

THE plants of the sea are phylogenetically very different from their terrestrial counterparts, and in open waters competitive advantage is with the small rather than with the large and complex as on land. Such differences justify treating marine botany as a distinct discipline but the task of surveying the whole field, including the taxonomy, life histories, biochemistry, physiology and ecology of marine plants, has not often been attempted. However, with the growth of our knowledge of the sea and the extension of our ability to see and handle marine plants *in situ* provided by SCUBA diving techniques, it is becoming easier to shed preconceived ideas derived from land-based botany and discern general principles.

Marine Botany is such a survey, designed chiefly for the undergraduate student. The author has set out to cover physiology and biochemistry, as well as taxonomic and ecological aspects, but the treatment is mainly descriptive. Thus a conventional account of the various groups of marine algae, of a sort which is available in many other texts, takes up almost half the book. The space left for other topics is not always used to best advantage — a picture of an oxygen meter is not very informative. The single chapter on plankton says nothing of one of its most fascinating aspects — the interplay of buoyancy control and exploitation of water movements which enables this type of organism to live a precarious existence between the darkness of the depths and nutrient starvation at the surface.

Rather better are the chapters on seagrass communities, mangrove swamps and coral reefs, which give accounts of communities not often featured in botanical texts, and that on economic utilization is an up-to-date summary. There are appendices on marine fungi and bacteria and these, too, contain material which is not readily available elsewhere. A feature of the book is that details of chemical and other techniques and keys for identification for some groups as far as order or genus are given. The author has written a book which will be useful as an introduction, but has rather missed the opportunity to display the beautiful adaptations in form and function of plants to the marine environment.

The Biology of Seaweeds, a multi-author work, is more limited in scope and much more advanced in treatment. It is organized in four sections — on structure and reproduction, ecology, physiology and biochemistry, and seaweeds as resources. The first of these summarizes recent advances in our knowledge of red, brown and green algae. For those already bemused by the intricacies of the life-cycles of red seaweeds these chapters will not help; but that physiologists, biochemists and ecologists must take account of such things is shown in an excellent account of polysaccharides — in which alternation of generations at the biochemical level is described — by the well-supported suggestion that having alternate life-forms better adapts a seaweed to environmental stress, and by casual observations such as that the crustose phase of *Scytosiphon lomentaria* is less palatable to the snail *Littorina littorea* than is the foliose plant.

Better understanding of life-cycles has arisen largely from increasing use of laboratory cultures, and cultures are also being used extensively in studies on nutrient uptake, morphogenesis, growth regulators and other biochemical and physiological aspects, while transplantation and other *in situ* manipulations are now standard techniques for the field ecologist. Paralleling this is the development of methods in mariculture to meet the growing demand for seaweed products in industry. Chapters on geographical distribution, populations, various aspects of seaweed photosynthesis, translocation, chemical constituents and sexuality are all authoritative and thought-provoking. A statement that a water movement of 14 m sec^{-1} exerts a stress equivalent to that of a gale of 890 miles per hour on land makes one wish that more had been said about the mechanical properties of seaweeds, in particular their relation to the molecular architecture of the cell wall.

However, the study of seaweeds is expanding too rapidly for any survey to be comprehensive. Apart from the proceedings of the ten international seaweed symposia which have now been held, this book is the first major work to bring together knowledge of the various aspects of these marine plants under one cover. It is an important sea-mark in marine botany that will soon be left astern. □

G.E. Fogg is Professor of Marine Biology at the University College of North Wales, Bangor.

A microbiologist's vade-mecum

John Postgate

The Prokaryotes: A Handbook on Habitats, Isolation, and Identification of Bacteria. Edited by Mortimer P. Starr *et al.* Two parts, pp.2,596, ISBN 3-540-08871-7. (Springer-Verlag: 1982.) DM 880, $410.

To do this magnificent work justice would require a team of reviewers almost as numerous as its 182 contributors. For it is nothing less than a comprehensive handbook of bacteriology: a detailed and up-to-date guide to the isolation, cultivation and identification of almost all the known bacteria.

The two volumes are divided into 23 sections (A–W inclusive), each of which comprises between 2 and 18 chapters. Section A consists of 6 introductory essays, the first of which is an excellent account of prokaryotic diversity which contrives to be highly informative as well as entertaining — see Table 13, for example. The other essays cover habitats, including diseased man and animals, as well as plants; principles of isolation, cultivation and conservation; and principles of characterization and identification. I found each of them thorough and informative; they provide an excellent *hors d'oeuvre* to the subsequent 22 courses.

I shall not have space to review those in any depth, so I shall indicate the pattern of the books by describing Section B, which deals with the phototrophic bacteria. It consists of 12 chapters. After a general introduction presenting their physiology and anatomy, come highly practical chapters on the isolation of cyanobacteria from normal and saline environments. A contribution on the habitats, isolation and characterization of vacuolate cyanobacteria is followed by one of broadly similar pattern on thermophilic cyanobacteria. Next is a more taxonomic contribution on the genera of cyanobacteria. The prochlorophytes are then briefly dealt with, preceding a short account of the isolation and cultivation of rhodospirilla. The genus *Ectothiorhodospira* gets its own chapter, followed by the chromatia and chlorobia; these two chapters collectively cover the phototrophic sulphur bacteria. The isolation of the photosynthetic flexibacteria is discussed next, and the section is concluded by a key to the classification of the non-oxygenic phototrophs. As this description indicates, the phototrophic bacteria are dealt with very thoroughly and the emphasis is generally highly practical: on media, culture, enrichment, storage and so on.

This pattern, a general essay introducing the detailed articles, is more or less followed in each section throughout the book. Section C is on the gliding bacteria, and D — the shortest, containing two chapters — deals with some sheathed and vacuolate bacteria. Section E covers budding and appendaged bacteria; F, the spirochaetes; G, some helical and curved bacteria; H, the pseudomonads; I, the diazotrophs with *Agrobacterium* and *Alcaligenes* thrown in. Section J is a rag-bag of mixed chemotrophs (sulphate reducers, methylotrophs, methanogens, halobacteria and so on), the least logical but nevertheless a most useful section. Section K covers obligate chemotrophs, bringing us to the end of what might broadly be called "environmental" bacteria.

It is a pity that Section L, the last in Vol.1, covering medically important Gram-negative rods (*Brucella, Legionella* and so on), was not transferred to the second volume but I suppose binding problems would have arisen. Volume 2 itself covers *inter alia* most of the medically important bacteria, as well as rumen and spoilage organisms. Section M is on enterobacteria and vibrios; N, various facultative rods; O, anaerobic rods; P and Q, cocci; R and S, spore formers. Section T, on coryneforms, leads logically to Section U on actinomycetes. Finally, Section V deals with the obligate symbionts and W with the wall-deficient bacteria.

I found it impracticable to go through all 169 chapters in detail but I have read carefully about 20 that I know something about and have dipped into most of the rest. As with the introductory essays, I am impressed by the quality and conciseness of the writing and the high information content. The authors are leading experts in their particular fields and their names, too many to list, will be familiar to all microbiologists. Some of the arrangement is idiosyncratic, however, and some of the classifications are inconsistent with the eighth edition of *Bergey's Manual* (Williams & Wilkins, 1974), usually for good reasons. No handbook of this kind can be perfect: I combed Section C (gliding bacteria) for a cross-reference to *Desulfonema,* which rightly appears among sulphate-reducing bacteria, although it glides. It is also a pity that all the colour pictures are clustered at the front of Vol.1.

A major concern to most potential customers will be the overlap with *Bergey's Manual.* Certainly, *The Prokaryotes* covers much of the information available in the eighth edition of *Bergey* but it is a far more ambitious and up-to-date work: it is not simply a determinative system, but a real laboratory handbook for microbiologists. I recommend it strongly to any microbiological library; it is an invaluable guide and reference work which ought to be readily available to every working microbiologist. □

John Postgate is Professor of Microbiology at the University of Sussex and Director of the Agricultural Research Council's Unit of Nitrogen Fixation.

Ubiquitous gels

E.G. Richards

Electrophoresis: Theory, Techniques, and Biochemical and Clinical Applications. By Anthony T. Andrews. Pp.336. ISBN 0-19-854626-2. (Clarendon/Oxford University Press: 1982.) £25, $59.

Dr Andrews remarks in his preface to this volume that well over half of all research papers in biochemistry involve electrophoresis. This impressive statistic calls for comment. Books with similar titles published 30 years ago would be mainly about free boundary electrophoresis à la Tiselius, a technique requiring complicated and expensive equipment; in contrast the present book is predominantly about zonal electrophoresis in polyacrylamide gels. For this the equipment is simple and cheap; the method is easy to master and no great sophistication is required to interpret the results. Anybody can do it, and, as the literature suggests, does.

Gels, polyacrylamide, starch or agarose provide the most favoured media for zonal electrophoresis with polyacrylamide well ahead of the field. It is 23 years since Raymond and Weintraub introduced the latter and thus strange that there is still no clear picture of the structure of such gels or how they exert their sieving effect in the electrophoretic process. Such understanding as we have is empirical. A consequence is that the technique still resembles more an art than a science. Recipes, with which the book abounds, are handed down from generation to generation of research students, and anomalies and artifacts, many of which Dr Andrews discusses, are there to dishearten the novice and mystify the expert.

Such theory that there is is adequately displayed, albeit occasionally a little shakily, and it is a minor criticism that it is developed in parallel for proteins and for nucleic acids. This leads to both repetition and a failure to indicate how lessons learnt from one class of macromolecules are relevant to the other.

Thus the major thrust is practical. Dr Andrews is to be congratulated on providing under one cover, succinct, clear and practical descriptions of electrophoresis in polyacrylamide and other media in most of the usual variations, as well as of related techniques including isoelectric-focusing, isotachophoresis, immuno-electrophoresis and affinity electrophoresis. □

E.G. Richards is Senior Lecturer at the Department of Biophysics at King's College, University of London.

BOOKS RECEIVED

Biological Sciences

COMPER, W. D. Heparin (and Related Polysaccharides). Polymer Monographs, Vol. 7. Pp.254. ISBN 0-677-05040-2. (Gordon & Breach: 1981.) $55.95.

CONRAT-FRAENKEL, H. and KIMBALL, P. C. Virology. Pp.406. ISBN 0-13-942144-0. (Prentice-Hall: 1982.) $32.95.

COOPER, S. J. (ed.). Theory in Psychopharmacology, Vol. 1. Pp.338. ISBN 0-12-18800-1. (Academic: 1981.) £19.20, $39.50.

DALE, J. E. The Growth of Leaves. Studies in Biology, 137. Pp.57. ISBN 0-7131-2836-4. (Edward Arnold: 1982.) £2.25.

EISLER, R. Trace Metal Concentrations in Marine Organisms. Pp.685. ISBN 0-08-025975-8. (Pergamon: 1981.) $120.

FELDBERG, W. S. Fifty Years On: Looking Back on Some Developments in Neurohumoral Physiology. Pp.106. ISBN 0-85323-364-0. (Liverpool University Press: 1982.) £6.25.

GALLAGHER, R. H. *et al.* (eds). Finite Elements in Biomechanics. Pp.401. ISBN 0-471-09996-1. (J. Wiley & Sons: 1982.) £21, $49.50.

GUTHRIE, W. and FAWKES, R. A Colour Atlas of Surgical Pathology. Pp.220. ISBN 0-7234-0759-2. (Wolfe Medical: 1982.) £35.

HALL, J.C., GREENSPAN, R.J. and HARRIS, W.A. Genetic Neurobiology. Pp.284. ISBN 0-262-08111-3. (MIT Press: 1982.) $29.95.

HAWK, G.L. (ed.). Biological/Biomedical Applications of Liquid Chromatography III. Chromatographic Science, Vol.18. Pp.440. ISBN 0-8247-1297-8. (Marcel Dekker: 1981.) SwFr. 148.

HAY, J.G. and REID, J.G. The Anatomical and Mechanical Bases of Human Motion. Pp.443. ISBN 0-13-035139-3. (Prentice-Hall: 1982.) £14.95.

HAYAT, M.A. Fixation for Electron Microscopy. Pp.501. ISBN 0-12-333920-0. (Academic: 1981.) $50.

HECHT, M. K., WALLACE, B. and PRANCE, G. T. (eds). Evolutionary Biology, Vol. 14. Pp.437. ISBN 0-306-40775-2. (Plenum: 1982.) $39.50.

HÉLÈNE, C. *et al.* (eds). Trends in Photobiology. Proc. of the 8th International Congress on Photobiology and of the Colloque International du CNRS held July 1980, Strasbourg, France. Pp.673. ISBN 0-306-40644-6. (Plenum: 1982.) $35.

HENLE, W. *et al.* (eds). Current Topics in Microbiology and Immunology, Vol. 94/95. Pp.308. ISBN 3-540-10803-3. (Springer-Verlag: 1981.) DM 132, $61.50.

HENLE, W. *et al.* (eds). Current Topics in Microbiology and Immunology, Vol.97. Pp.204. ISBN 3-540-11118-2. (Springer-Verlag: 1982.) DM 94. $43.80.

HENRICH, H. (ed.). Microvascular Aspects of Spontaneous Hypertension. Pp.186. ISBN 3-456-81135-7. (Hans Huber: 1982.) SwFr.53, DM58.

HEYM, C. and FORSSMANN, W.-G. (eds). Techniques in Neuroanatomical Research. Pp.395. ISBN 3-540-10686-3. (Springer-Verlag: 1981.) DM 118, $55.

HOERNI-SIMON, G. and ZITTOUN, R. (coordinateurs) Progrès en Hématologie, 2. Décembre 1981. Les Lymphomes Malins Non Hodgkiniens. Pp.166. Pbk ISBN 2-7040-0406-4. (Doin Éditeurs, Paris: 1981.) Pbk np.

HOLZER, H. (ed.). Metabolic Interconversion of Enzymes, 1980. International Titisee Conference, October 1980. Pp.212. ISBN 3-540-10979-X. (Springer-Verlag: 1981.) DM 129, $60.10.

HOOK, E.B. and PORTER, I.H. (eds). Population and Biological Aspects of Human Mutation. Pp.435. ISBN 0-12-355440-3. (Academic: 1981.) $34.50.

HOPPENSTEADT, F. Mathematical Methods of Population Biology. Cambridge Studies in Mathematical Biology, 4. Pp.149. Hbk ISBN 0-521-23846-3; pbk ISBN 0-521-28256-X. (Cambridge University Press: 1982.) Hbk £15; £6.50.

HORECKER, B.L. and STADTMAN, E.R. (eds). Current Topics in Cellular Regulation, Vol.20 — 1981. Pp.317. ISBN 0-12-152820-0. (Academic: 1981.) $39.

HRABA, T. and HAŠEK, M. (eds). Cellular and Molecular Mechanisms of Immunologic Tolerance. Pp.600. ISBN 0-8247-1552-7. (Marcel Dekker: 1981.) SwFr. 178.

ILMAVIRTA, V., JONES, R.I. and PERSSON, P.-E. (eds). Lakes and Water Management. Developments in Hydrobiology, 7. Proceedings of the 30 Years Jubilee Symposium of the Finnish Limnological Society, held in Helsinki, Finland, September 1980. Pp.222. ISBN 90-6193-758-2. (W. Junk, The Hague: 1982.) Dfl.125, $54.50.

INGLE, D.J., GOODALE, M.A. and MANSFIELD, R.J.W. (eds). Analysis of Visual Behavior. Pp.434. ISBN 0-262-09022-8. (MIT Press: 1982.) $75.

JAKOBY, W.B. (ed.). Methods in Enzymology. Vol.77, Detoxication and Drug Metabolism: Conjugation and Related Systems. Pp.476. ISBN 0-12-181977-9. (Academic: 1981.) $45.

JECHT, E. W. and ZEITLER, E. (eds). Varicocele and Male Infertility. Recent Advances in Diagnosis and Therapy. Pp.211. ISBN 3-540-10727-4/0-387-10727-4. (Springer-Verlag: 1982.) DM88, $45.70.

KIRPICHNIKOV, V.S. Genetic Bases of Fish Selection. Translated from the Russian by G.G. Gause. Pp.410. ISBN 3-540-10911-0. (Springer-Verlag: 1981.) DM 98, $45.70.

KRIEGER, D.T. Cushing's Synodrome, Vol.22. Pp.142. ISBN 3-540-10811-4/0-387-10811-4. (Springer-Verlag: 1982.) DM88, $41.

LAIDLER, E. Otters in Britain. Pp.200. ISBN 0-7153-8069-9. (David & Charles: 1982.) £7.95.

LANGER, R. H. M. and HILL, G. D. Agricultural Plants. Pp.310. Hbk ISBN 0-521-22450-0; pbk ISBN 0-521-29506-8. (Cambridge University Press: 1982.) Hbk £20; pbk £7.95.

LARWOOD, G.P. and NIELSEN, C. Recent and Fossil Bryozoa. Papers Presented at the 5th International Conference, Durham 1980. Pp.334. Pbk ISBN 87-85215-11-2. (Olsen & Olsen, Fredensborg, Denmark: 1981.) Pbk $55.

LESSELL, S. and VAN DALEN, J. T. W. Neuro-ophthalmology 1982, Vol. 2. Pp.404. ISBN 90-219-3059-5/0-444-90210-4. (Excerpta Medica/Elsevier: 1982.) Dfl.190, $88.25.

LEWIS, D. H. (ed.). Induced Skeletal Muscle Ischemia in Man. Symposium held November 6-7, 1980, Linkoping. Pp.180. ISBN 3-8055-3427-2. (Karger, Basel: 1982.) SwFr.115, DM138, $69.

LOTE, C. J. Principles of Renal Physiology. Pp.179. Hbk ISBN 0-7099-0078-3; pbk ISBN 0-0799-0079-1. (Croom Helm: 1982.) Hbk £12.95; pbk £6.95.

THE MACDONALD ENCYCLOPEDIA OF TREES. Pp.300. Pbk ISBN 0-356-08574-0. (Macdonald, London: 1982.) Pbk £4.95.

MARAMOROSCH, K. and RAYCHAUDHURI, S.P. (eds). Mycoplasma Diseases of Trees and Shrubs. Pp.362. ISBN 0-12-470220-1. (Academic: 1981.) $29.50.

MAYERSDORF, A. and SCHMIDT, R.P. (eds). Secondary Epileptogenesis. Pp.188. ISBN 0-89004-578-X. (Raven: 1982.) $21.

MÉNACHÉ, D., SURGENOR, D. MacN. and ANDERSON, H.D. (eds). Hemophilia and Hemostatis. American Red Cross 13th Annual Scientific Symposium, Washington, DC, May 1981. Progress in Clinical and Biological Research, Vol.72. Pp.282. ISBN 0-8451-0072-9. (Alan R. Liss, New York: 1981.) £20.50, DM 90.

MERRELL, D.J. Ecological Genetics. Pp.500. ISBN 0-582-46349-1. (Longman: 1982.) £15.

MITCHELL, A. and WILKINSON, J. The Trees of Britain and Europe. Pp.264. Hbk ISBN 0-00-219037-0; pbk ISBN 0-00-219035-4. (Collins: 1982.) Hbk £6.95; pbk £3.95.

MOYLE, P.B. and CECH, J.J. Jr. Fishes: An Introduction to Ichthyology. Pp.293. ISBN 0-13-319723-9. (Prentice-Hall: 1982.) $29.95.

MYANT, M.B. The Biology of Cholesterol and Related Steroids. Pp.910. ISBN 0-433-22880-6. (Heinemann, London: 1981.) £45.

NATORI, S., IKEKAWA, N. and SUZUKI, M. (eds). Advances in Natural Products Chemistry. Extraction and Isolation of Biologically Active Compounds. Pp.599. ISBN 0-470-27245-7. (Wiley: 1981.) £66, $120.

NESTER, E.W. *et al.* The Microbial Perspective. Pp.730. ISBN 0-03-047041-2. (Saunders College Publishing: 1982.) Np.

NOBILE, S. and WOODHILL, J.M. Vitamin C. The Mysterious Redox-System. A Trigger of Life? Pp.185. ISBN 0-85200-419-2. (MTP Press, Lancaster: 1981.) £11.95.

ODDIE, B. Bill Oddie's Little Black Bird Book. Pp.148. Pbk ISBN 0-413-49480-2. (Methuen: 1982.) Pbk £1.50.

ORR, R.T. Vertebrate Biology. 5th Edn. Pp.568. ISBN 0-03-057959-7. (Saunders: 1982.) Np.

ORTNER, D.J. and PUTSCHAR, G.J. Identification of Pathological Conditions in Human Skeletal Remains. Smithsonian Contributions to Anthropology, No.28. Pp.479. (no ISBN) (Smithsonian Institution: 1981.) Pbk np.

PARKER, T.A. III, ALLEN PARKER, S. and PLENGE, M.A. An Annotated Checklist of Peruvian Birds. Pp.104. ISBN 0-931130-07-7. (Buteo: 1982.) $15.

PATTERSON, B. and WOOD, A.E. Rodents from the Deseadan Oligocene of Bolivia and the Relationships of the Caviomorpha. Bulletin of the Museum of Comparative Zoology, Vol.149: No.7. 12 January 1982. Pp.172. Pbk ISBN 0027-4100. (Harvard University: 1982.) Pbk $3.25.

PENNY, N.D. and ARIAS, J.R. Insects of an Amazon Forest. Pp.269. ISBN 0-231-05266-9. (Columbia University Press: 1982.) $48.75.

PESTKA, S. (ed.). Methods in Enzymology. Vol. 79. Interferons, Pt B. Pp.677. ISBN 0-12-181979-5. (Academic: 1981.) $59.

PICK, E. (ed.). Lymphokines. A Forum for Immunoregulatory Cell Products, Vol. 3. Pp.450. ISBN 0-12-432003-1. (Academic: 1981.) Np.

PODESTA, R. B. (ed.). Membrane Physiology of Invertebrates. Pp.664. ISBN 0-8247-1503-9. (Marcel Dekker: 1982.) Np.

POISSONET, P. *et al.* (eds). Vegetation Dynamics in Grasslands, Heathlands and Mediterranean Ligneous Formations. Advances in Vegetation Science, 4. Symposium of the Working Groups for Succession Research on Permanent Plots, and Data-processing in Phytosociology held at Montpellier, France, September 1980. Pp.286. ISBN 90-6193-636-5. (W. Junk, The Hague: 1982.) Dfl.195, $85.

POMEROY, L.R. and WIEGERT, R.G. (eds). The Ecology of a Salt Marsh. Ecological Studies — Analysis and Synthesis, Vol.38. Pp.271. ISBN 3-540-90555-3. (Springer-Verlag: 1981.) DM 69, $32.20.

RANSOM, R. (ed.). A Handbook of *Drosophila* Development. Pp.289. Hbk ISBN 0-444-80366-1; pbk ISBN 0-444-80418-8. (Elsevier: 1982.) Hbk Dfl.250, $116; pbk Dfl.100, $46.

RATTAZZI, M.C., SCANDALIOS, J.G. and WHITE, G.S. (eds). Isozymes. Current Topics in Biological and Medical Research, Vol.5. Pp.240. ISBN 0-8451-0254-0. (Alan R. Liss, New York: 1981.) £24.60, DM 108.

RAY, D.S. (ed.) The Initiation of DNA Replication. ICN–UCLA Symposia on Molecular and Cellular Biology, Vol.22, 1981. Proceedings of the 1981 ICN–UCLA Symposia held at Salt Lake City, Utah, March, 1981. Pp.628. ISBN 0-12-583580-9. (Academic: 1981.) $53.

RISSER, P.G. *et al.* The True Prairie Ecosystem. Pp.557. ISBN 0-87933-361-8. (Academic: 1981.) $31.50.

ROLLAND, R. *et al.* (eds). Follicular Maturation and Ovulation. Proceedings of the 4th Reinier de Graaf Symposium, August 20-22, 1981, Nijmegen. Pp.432. ISBN 90-219-0505-1/0-444-90231-7. (Excerpta Medica/Elsevier: 1982.) Dfl.195, $90.75.

RUBIN, R.H. and YOUNG, L.S. (eds). Clinical Approach to Infection in the Compromised Host. Pp.653. ISBN 0-306-40679-9. (Plenum: 1981.) $59.50.

RUTHERFORD, G.K. (ed.). The Physical Environment of the Faeroe Islands. Monographiae Biologicae, Vol.46. Pp.147. ISBN 90-6193-099-5. (W. Junk, The Hague: 1982.) Dfl.95, $39.50.

SANDERS, G. L. *et al.* (eds). Pulmonary Toxicology of Respirable Particles. Proc. of the 19th Annual Hanford Life Sciences Symposium, October 22-24, 1979, Richland. Pp.676. ISBN 0-87079-121-4. (US Department of Energy: 1980.) $25.25.

SAN PIETRO, A. (ed.). Biosaline Research: A Look to the Future. Environmental Science Research, Vol.23. Proceedings of the 2nd International Workshop held in La Paz, Mexico, November 1980. Pp.578. ISBN 0-306-40892-9. (Plenum: 1982.) $65.

SATO, G. (ed.). Functionally Differentiated Cell Lines. Pp.280. ISBN 0-8451-0241-1. (Alan R. Liss, New York: 1981.) £25.90, DM 114.

SEGRE, D. and SMITH, L. (eds). Immunological Aspects of Aging. Immunology Series, Vol.15. Pp.584. ISBN 0-8247-1349-4. (Marcel Dekker: 1981.) SwFr. 178.

SILVERSTONE, T. (ed.). Drugs and Appetite. Pp.200. ISBN 0-12-643780-7. (Academic: 1982.) £14.20, $29.50.

SMITH, H. (ed.). Plant and the Daylight Spectrum. Pp.508. ISBN 0-12-650980-8. (Academic: 1981.) £22, $45.50.

SPENCE, A. P. Basic Human Anatomy. Pp.704. ISBN 0-8053-6994-5. (Benjamin/Cummings: 1982.) $27.95.

STAMATOYANNOPOULOS, G. and NIENHUIS, A.W. (eds). Hemoglobins in Development and Differentiation. Pp.526. ISBN 0-8451-0213-3. (Alan R. Liss, New York: 1981.) Np.

TU, A. T. (ed.). Survey of Contemporary Toxicology, Vol. 2. Pp.248. ISBN 0-471-06352-5. (Wiley: 1982.) Np.

WEXLER, P. Information Resources in Toxicology. Pp.291. ISBN 0-444-00616-8. (Elsevier/North-Holland: 1982.) Np.

WOOLHOUSE, H. W. (ed.). Advances in Botanical Research, Vol. 9. Pp.279. ISBN 0-12-005909-6. (Academic: 1981.) £24, $49.50.

Applied Biological Sciences

ANDREOLI, M., MONACO, F. and ROBBINS, J. (eds). Advances in Thyroid Neoplasmia 1981. Pp.363. ISBN 88-7084-000-X. (Acta Medica: 1981.) L10,000.

BARNETT, H. J. M., HIRSH, J. and FRASER MUSTARD, J. (eds). Acetylsalicylic Acid: New Uses for an Old Drug. Pp.288. ISBN 0-89004-647-6. (Raven: 1982.) $53.04.

BLUME, W. T. Atlas of Pediatrics Electroencephalography. Pp.344. ISBN 0-89004-564-X. (Raven: 1982.) $85.

BOZLER, G. and VAN ROSSUM, J. M. (eds). Pharmacokinetics During Drug Development: Data Analysis and Evaluation Techniques. Drug Development and Evaluation, Vol.6. Pp.328. ISBN 3-437-10654-6. (Gustav Fischer: 1982.) DM 98.

CARPENTER, D. C. (ed.). Cellular Pacemakers. Mechanisms of Pacemaker Generation, Vol. 1. Pp.332. ISBN 0-471-06509-9. (Wiley: 1982.) Np.

DE SERRES, F. J. and HOLLAENDER, A. (eds). Chemical Mutagens. Principles and Methods for their Detection, Vol. 7. Pp.479. ISBN 0-306-40771-X. (Plenum: 1982.) $49.50.

FIECHTER, A. (ed.). Advances in Biochemical Engineering. Vol. 21, Microbes and Engineering Aspects. Pp.232. ISBN 3-540-11019-4. (Springer-Verlag: 1982.) DM92, $42.90.

GEORGIEVSKII, V.I., ANNANKOV and SAMOKHIN, V.T. Mineral Nutrition of Animals. Studies in the Agricultural and Food Sciences. Pp.475. ISBN 0-408-10770-7. (Butterworths: 1981.) £45.

HÄMMERLING, G. J., HÄMMERLING, U. and KEARNEY, J. F. (eds). Monoclonal Antibodies and T-Cell Hybridomas: Perspectives and Technical Advances. Research Monographs in Immunology, Vol. 3. Pp.587. ISBN 0-444-80351-3. (Elsevier Biomedical: 1981.) $99.50, Dfl.234.

HAY, J. B. (ed.). Animal Models of Immunological Processes. Pp.260. ISBN 0-12-333520. (Academic: 1982.) £21.20, $43.50.

HOFSCHNEIDER, P. H. and GOEBEL, W. (eds). Gene Cloning in Organisms Other Than *E. coli*. Current Topics in Microbiology and Immunology, Vol.96. Pp.264. ISBN 3-540-11117-4. (Springer-Verlag: 1982.) DM92, $42.90.

HOLLAENDER, A. *et al.* (eds). Genetic Engineering of Microorganisms for Chemicals. Basic Life Sciences, Vol.19. Proceedings of a Symposium held May 1981, at the University of Illinois at Champaign-Urbana. Pp.485. ISBN 0-306-40912-7. (Plenum: 1982.) $55.

HOSKING, G. An Introduction to Paediatric Neurology. Pp.252. Hbk ISBN 0-571-11848-8; pbk ISBN 0-571-11849-6. (Faber & Faber, London: 1982.) Hbk £11.50; pbk £5.95.

HSU, J. M. and DAVIS, R. L. (eds). Handbook of Geriatric Nutrition: Principles and Applications for Nutrition and Diet in Aging. Pp.372. ISBN 0-8155-0880-8. (Noyes, New Jersey: 1981.) $39.

HSU, T. C. (ed.). Cytogenetic Assays of Environmental Mutagens. Pp.412. ISBN 0-916672-56-5. (Allanheld Osmun: 1982.) $45.

LOWRANCE, W. W. (ed.). Assessment of Health Effects at Chemical Disposal Sites. Proceedings of a Symposium held June 1–2, 1981, New York City. Pp.162. ISBN 0-86576-025-X. (William Kaufmann: 1981.) $10.

RHODES-ROBERTS, M. and SKINNER, F. A. (eds). Bacteria and Plants. Society for Applied Bacteriology Symposium, 10. Pp.264. ISBN 0-12-587080-9. (Academic: 1982.) £14.50, $29.50.

Psychology

BRANDSTATTER, H., DAVIS, J. H. and STOCKER-KREICHGAUER, G. (eds). Group Decision Making. European Monographs in Social Psychology, 25. Pp.500. ISBN 0-12-125820-3. (Academic: 1982.) £21.80, $45.

HOFFMEISTER, F. and STILLE, G. (eds). Psychotropic Agents. Part II, Anxiolytics, Gerontopsychopharmacological Agents, and Psychomotor Stimulants. Pp.778. ISBN 3-540-10300-7/0-387-10300-7. (Springer-Verlag: 1981.) DM 360, $134.10.

KEMPTON, W. The Folk Classification of Ceramics. A Study of Cognitive Prototypes. Pp. 237. ISBN 0-12-404080.2. (Academic: 1981.) $24.50.

MAIN, C. J. (ed.). Clinical Psychology and Medicine: A Behavioural Perspective. Pp.361. ISBN 0-306-40900-3. (Plenum: 1982.) $39.50.

SAHAKIAN, W. S. (ed.). History and Systems of Social Psychology, 2nd Edn. Pp.656. Hbk ISBN 0-89116-179-1; pbk ISBN 0-89116-191-0. (Hemisphere: 1982.) Hbk $24.50, pbk $16.95.

SHEPHERD, M. *et al.* Psychiatric Illness in General Practice, 2nd Edn. Pp.238. ISBN 0-19-261243-3. (Oxford University Press: 1981.) £15.

Anthropology

BAILLIE, M.G.L. Tree-Ring Dating and Archaeology. Pp.274. ISBN 0-7099-0613-7. (Croom Helm, London: 1982.) £16.95.

BROTHWELL, D.R. Digging up Bones: The Excavation, Treatment and Study of Human Skeletal Remains. 3rd Edn. Pp.208. Hbk ISBN 0-19-858504-7; pbk ISBN 0-19-858510-1. (British Museum (Natural History)/Oxford University Press: 1982.) Hbk np; pbk £8.95.

JONES, G.D. and KAUTZ, R.R. (eds). The Transition to Statehood in the New World: New Directions in Archaeology. Pp.254. ISBN 0-521-24075-1. (Cambridge University Press: 1981.) £17.50.

MAI, L.L., SHANKLIN, E. and SUSSMAN, R.W. (eds). The Perception of Evolution. Essays Honoring Joseph B. Birdsell. Anthropology UCLA, Vol. 7, Numbers 1 & 2, 1981. Pp.273. (no ISBN) (University of California, Los Angeles: 1981.) Pbk np.

General

ABDULLAH, T.A. and ZEIDENSTEIN, S.A. Village Women of Bangladesh: Prospects for Change. Women in Development Series, Vol.4. Pp.246. ISBN 0-08-026795-5. (Pergamon: 1981.) £10, $23.

ALLAN, P. and JOLLEY, M. (eds). Nursing, Midwifery & Health Visiting Since 1900. Pp.316. Hbk ISBN 0-571-11839-9; pbk ISBN 0-571-11840-2. (Faber & Faber, London: 1982.) Hbk £7.95; pbk £4.95.

ANDORKA, R. Determinants of Fertility in Advanced Societies. Pp.431. Pbk ISBN 0-416-67350-3. (Methuen: 1982.) Pbk £5.95.

BARNABY, F. and THOMAS, G. The Nuclear Arms Race–Control or Catastrophe? Pp.250. ISBN 0-86187-229-0. (Francis Pinter (Publishers) Ltd., 5 Dryden Street, London WC2E 9NW: 1982.) £12.50.

BARNES, B. T. S. Kuhn and Social Science. Pp.126. Hbk ISBN 0-333-28936-6; pbk ISBN 0-333-28937-4. (Macmillan, London: 1982.) Hbk £12; pbk £3.95.

BERRY, G. A Tale of Two Lakes. The Fight to Save Ennerdale Water and Wastwater. Pp.96. Pbk ISBN 0-9504629-2-6. (Friends of the Lake District, Kendal, Cumbria: 1982.) Pbk £2.25.

BHAT, D.H.S. The Sixth Sense. Pp.105. (no ISBN) (Bharati Bhat, India: 1982.) Rs 30, $5.

BLACK, M. and REED, J. S. (eds). Perspectives on the American South. Annual Review of Society, Politics and Culture, Vol. 1. Pp.409. ISBN 0-677-16260-X. (Gordon & Breach: 1981.) $35.

BLEECKER, D. Gauge Theory and Variational Principles: Global Analysis. Pp.180. ISBN 0-201-10096-7. (Addison-Wesley: 1981.) $17.50.

BOYDEN, S., MILLAR, S., NEWCOMBE, K. and O'NEILL, B. The Ecology of a City and its People: The Case of Hong Kong. Pp.427. ISBN 0-7081-1095-9. (Australian National University Press: 1981.) $A17.50.

BRINKMANN, K. and SCHAEFER, H. (eds). Der Klektrounfall. Pp.324. ISBN 3-540-11003-8/0-387-11003-8. (Springer-Verlag: 1982.) DM 128, $59.60.

COX, R.A.F. (ed.). Offshore Medicine: Medical Care of Employees in the Offshore Oil Industry. Pp.199. ISBN 3-540-11111-5; 0-387-11111-5. (Springer-Verlag: 1982.) DM 80, £19.

FINDLEY, D.F. (ed.). Applied Time Series II. Proceedings of the 2nd Symposium held in Tulsa, Oklahoma, March 1980. Pp.798. ISBN 0-12-256420-0. (Academic: 1981.) $49.50.

FISHER, A.C. Resource and Environmental Economics. Pp.284. Hbk ISBN 0-521-24306-8; pbk ISBN 0-521-28594-1. (Cambridge University Press: 1982.) Hbk £20; pbk £6.95.

INSTITUTE OF SOIL SCIENCE, ACADEMIA SINICA. (eds). Proceedings of Symposium on Paddy Soils, Beijing, The People's Republic of China. Pp.864. ISBN 3-540-10900-5. (Springer-Verlag: 1981.) DM 98, $45.70.

KERR, F.W.L. The Pain Book. Pp.192. Hbk ISBN 0-13-647826-3; pbk ISBN 0-13-647818-2. (Prentice-Hall: 1981.) Np.

LAUDAN, L. Science and Hypothesis: Historical Essays on Scientific Methodology. Pp. 240. Hbk ISBN 90-277-1315-4; pbk ISBN 90-277-1316-2. (Reidel: 1981.) Hbk Dfl. 80, $34.95; pbk Dfl. 37.50, $14.95.

MACKLIN, R. Man, Mind, and Morality. The Ethics of Behavior Control. Pp.130. Pbk ISBN 0-13-551127-5. (Prentice-Hall: 1982.) Pbk £5.65.

MALCOLM, A. and POYSER, J. (eds). Computers and the General Practitioner. Pp.124. ISBN 0-08-026865-X. (Pergamon: 1981.) £10, $20.

MANN, W.B. Was There a Fifth Man? Quintessential Recollections. Pp.170. ISBN 0-08-027445-5. (Pergamon: 1982.) £9.50, $19.50.

NISBET, R.M. and GURNEY, W.S.C. Modelling Fluctuating Populations. Pp.379. ISBN 0-471-28058-5. (Wiley: 1982.) £19.50, $47.50.

OMRAN, A.R. and STANDLEY, C.C. (eds). Family Formation Patterns and Health. Further Studies: An International Collaborative Study in Columbia, Egypt, Pakistan and the Syrian Arab Republic. Pp.464. Pbk ISBN 92-4-156070-3. (WHO: 1981.) Pbk SwFr. 44.

PRYDE, E.H., PRINCEN, L.H. and MUKHERJEE, K.D. (eds). New Sources of Fats and Oils. Symposium at the International Society for Fat Research/American Oil Chemists' World Congress, April/May 1980. Pp.340. (no ISBN) (American Oil Chemists' Society, Champaign, Illinois: 1981.) Np.

RENFREW, C. and WAGSTAFF, M. (eds). An Island Polity. The Archaeology of Exploitation in Melos. Pp.361. ISBN 0-521-23785-8. (Cambridge University Press: 1982.) Np.

STOLLER, B. B. The Birth of the Imagination. Pp.341. (No ISBN) (Stoller Research Co., California: 1982.) Np.

YIN, R. K. Conserving America's Neighborhoods. Pp.195. ISBN 0-306-40795-7. (Plenum: 1982.) $25.

nature

24 June 1982

Making a strategy for Europe

The European Commission's heart may be in the right place in the support of science and technology, but it must learn patience. Europe will be there a century from now.

The European Community is going through a strange and unrecognized metamorphosis. President Ronald Reagan has come and gone, leaving behind him a trail of disappointment. What worries Europeans most is not that he was less than well-informed about their affairs but that so little happened or got done. The Falklands war has been settled, after a fashion, reminding Europeans that in the last resort blood, which means financial common interest, is thicker than water. However reluctantly, the members of the Community found that they had no choice but to go along with the unexpected British assertion of national purpose (and Irish diffidence on this point so to speak is the exception that proves the rule). Meanwhile, the whole of Europe is perplexed that there should be little it can do about the mounting crisis in the Middle East. There is almost nothing that can be done to avert the slow erosion of Chancellor Helmut Schmidt's government in Bonn. This and the concatenation of dismal happenings has, however, had a paradoxical effect — Europeans seem to have been convinced that they are in the same boat, and that they had better make the best of it. The constructive performance of the government of France at the Versailles summit, however eccentric, has been heartening for everybody on the east of the Atlantic. This, strange as it may seem, is the spirit in which the meeting of the council of science ministers of the Community this week should be judged.

The objective of the meeting is to make a further step towards a science policy for the European Community. It tends to be forgotten that the Community has had a science policy since before its own creation — a policy called Euratom. It is, however, well within living memory that whenever the Community has since tried to hammer out some concerted plan of action on its support of science and technology, the result has been conspicuously inconspicuous. Plans for coordinated programmes of research and development in computers, or telecommunications, have regularly foundered or have been still-born. The explanation is familiar but unremembered — national governments will not commit national funds to a common programme of research and development if they fear that their gain will be proportionately less than their financial contribution. To ensure that such an outcome will not arise, they insist that whatever cooperative programmes there may be should be tightly controlled by representatives of the Community's member states, which is a recipe for making sure that the programmes will be second-rate. Will this week's meeting have a better chance of succeeding?

This meeting (see page 619) is at least well prepared. The council of ministers of science has been deluged with paper. There is also a hefty background document to direct attention to the directions in which, on a kind of statistical-historical view of events, the wind is blowing (see *Nature* 17 June, p.528). The Commission's new proposals show that it has learned something from the disappointments of the past quarter of a century; the intention now is that there should be a political decision about the fields in which Community funds might be spent in the interests of the Community as a whole, a kind of political assessment of more detailed plans for spending money in the selected areas and then, finally, the making of research grants by specialist committees. The procedure proposed has the obvious advantage of putting a distance between the politicians who provide the funds and those who ultimately decide how they should be spent. The distance, however, is unlikely to be great. The committees that spend the money will be creatures of the member governments, and their members not so independent that they will be able to snap their fingers at those who nominate them. The best to hope for is that the regime now proposed will repeat past mistakes less glaringly.

Before it is too late, there is another model that European governments should consider — the agreement signed last week between the principal research councils of the United Kingdom and the Netherlands (see page 616). Even this marriage of convenience has arisen more or less by accident, partly because the British and the Dutch share common scientific interests and frustrations and partly because they get on well with each other in a wider context. Moreover, the deal now struck by the research councils falls far short of being a full marriage — there is no suggestion that British funds should for the time being be used to support research carried out in Dutch laboratories by nationals of the Netherlands. Yet the benefits of even this arrangement are classically simple. In at least some respects, the community of ideas will be enlarged.

The lesson for the European Community in this tale is that it takes time — and decades, not years — to bring about a tangible sense of collaboration between naturally competitive creative people, and that the collaboration works well only when the people who decide what should be done are the people who will then do the work. The model is persuasive, but by no stretch of the imagination quick enough for the European Commission's purpose. How, the Commission will be asking, can the "technology gap" then be bridged?

In reality, there must be some middle road. For political reasons, the Commission seeks to invest in research, necessarily applied research, that will both strengthen the sense of community and yield some practical benefit in the short run. Notoriously, what it wants can be accomplished only in the sphere in which commercial companies are strong and even dominant. It would be better advised to back away from the goal of immediate benefit, seeking instead to improve the cohesion of the research community. The imaginative use of Community funds for helping people from member states to move more freely within the Community would be a start, especially if the scope of such a scheme could be enlarged beyond the fields now cultivated by the European academies and by the European Science Foundation. Modest grant-making through independent bodies of trustee-like dignatories would help still further if the grants found their way to those who in the judgement of their peers deserved them. Such painstaking work cannot, of course, be counted on to close the technology gap by some arbitrarily chosen target date, but no deliberate policy will do that (although luck might). But if Europe is here to stay, as the events of the past few weeks appear to attest, it will still be here a century or so from now.

US contributions

Authors in the United States and Canada are reminded that they should send their manuscripts either direct to London or to the Washington office at 991 National Press Building, Washington DC20045; manuscripts should *not* be sent to *Nature*'s New York office.

0028-0836/82/250613 02$01.00

Antarctic trouble ahead?

The welcome end of the Falklands war may yet herald trouble for the Antarctic Treaty.

The shooting may have stopped in the Falkland Islands but the state of war continues. Britain says that the exclusion zone in which Argentine ships will be regarded as hostile will remain in place around the Falklands until it is assured that hostilities are at an end. But the government of Argentina, in the throes of its post-war reconstruction, cannot yet make up its mind what should be done. Whatever government emerges in Buenos Aires, however, the Argentine claim of sovereignty over the Falklands is unlikely simply to be abandoned. The flat contradiction of that cry by the British government's insistence that the Falklands are a British possession — reinforced by the removal of an Argentine party from Thule in the South Sandwich Islands at the weekend — is a depressing omen for the future. The best chance that civility can return to Anglo-Argentine relations is that the issue of sovereignty should be put on ice, but the diplomats will be hard pressed to devise an agreement to that effect which each government can stomach. The prospect is that the negotiations that eventually take place will be as protracted and frustrating as those which preceded the war — and will be embittered because of it. The hope must be that this souring of relations will not spill over into the delicate negotiations now taking place under the umbrella of the Antarctic Treaty.

Britain and Argentina are key players in the circle of 14 nations that govern the region south of 60 degrees south latitude through the treaty, which came into effect in 1961. In the past few years, this group, building on a long history of mutual cooperation in the Antarctic, has begun a chain of decisions about how to dispose of the Antarctic region's food and mineral resources. Two critical meetings on these topics have taken place this month and it is a measure of the importance all parties attach to the subject that the meetings went forward despite the Falklands war.

The earlier meeting, at Hobart in Tasmania, was the first of the new commission established under the 1980 convention for the conservation of Antarctic marine living resources. Britain's scientific contribution, particularly that of Richard M. Laws, director of the British Antarctic Survey, has been important in establishing the concept that there is a serious risk that Antarctic krill — the region's principal food resource — could be overfished. Indeed, British research has shown that overfishing of some species of Antarctic fish has already taken place. Overfishing of krill (probably by the Soviet Union and Japan, which took perhaps as much as 500,000 tonnes last season) could imperil the higher life forms of the Antarctic that depend on krill as their main food source. Most endangered would be Antarctic whale stocks, already depleted by whaling earlier in the century. Formally, the commission could crack down on countries accused of overfishing, but the meeting in Hobart was also intended to establish a new scientific committee. This group is crucial to the success of the new convention, for it is the technical link that will determine when krill are being overfished. Since British research has contributed so fully to the understanding of Antarctic marine resources, it had been expected that Laws would become its first chairman. The fact that his election has now been vetoed is a clear sign of fall-out from the Falklands, and a warning that the amity of the treaty is in danger.

The second meeting, in Wellington, New Zealand, has been planned to make a start on the thorny question of the ultimate disposition of Antarctic minerals. How can exploitation be made compatible with the original treaty? The continent has huge continental shelves (the Weddell Sea floor is the size of Venezuela) with thick sedimentary deposits that may contain hydrocarbon deposits for example. One snag is that nobody agrees which country owns which part of Antarctica. Indeed, Britain, Argentina and Chile have overlapping and conflicting claims to the part south of South America, including the Weddell Sea. Whether it will be possible to allocate rights to mineral resources before the ownership of the land has been settled is problematical, but there is no shortage of technical questions to be hammered out in advance — the technical steps that must be taken to avoid pollution of this unique environment for example.

The most obvious danger is that with the parties to the Falklands dispute also prominent signatories of the treaty, mischief-making will break out. Of claimants on the Antarctic, Argentina is among the most clamant, and has even sought to reinforce its claims by sending pregnant women to Antarctica so that their babies can be born Antarctic Argentines. Hitherto, Britain has been by comparison a modest claimant, saying that its formal claims in Antarctica are intended more as declarations of trusteeship than as old-fashioned territorial claims. It is to be hoped that the experience of the past three months will not have given the British too keen an appreciation of the pleasures of possession. Either way, since all decisions within the Antarctic Treaty must be taken unanimously, it will be easy for Argentina, Britain or some third party effectively to veto an unwelcome proposal by saying that the question cannot be settled until ownership has been resolved. By the convention signed in 1980, it has been agreed that something should be done to exploit marine resources sensibly, but nothing has been decided about minerals.

The meeting in Wellington has been made necessary by the need to forestall unwelcome activities in the Antarctic. The most obvious danger is that a failure to agree a set of rules will let in a host of cowboys. The best but also the most likely outcome is that there should be a moratorium on mineral exploitation until procedures have been worked out for telling how the profits (if any) should be distributed. But it should be fresh in the minds of member states that a very similar question occupied the conference on the Law of the Sea for the best part of a decade. It is not too soon to start on this negotiation.

Ships will go down

The Falklands conflict seems to have heartened the British naval lobby. It should not.

After its successful reoccupation of the Falkland Islands, the British government and especially the defence department finds itself engaged in a war that could politically or financially more damaging — a war with the most loyal of its supporters in the past three months, the members of the Navy lobby. Like the Falklands war, the war about naval policy has been rumbling away for months and even years, on one reckoning since in the 1960s the British government gave up the pretence of keeping a strong strategic force east of Suez and stopped building large aircraft carriers soon afterwards. More recently, however, the present government has had to forgo the services of one Navy minister (Mr Michael Steed) who resigned because he thought the British Navy was being starved of funds (Trident nuclear submarines not counting) while Mr John Nott, the present defence minister, appears to have succeeded Mr Francis Pym (by accident, now Foreign Secretary) because he showed more willingness to tell the Navy that they cannot have all the equipment they would like. Inevitably, in the weeks following a successful naval engagement, the case for building more ships is being loudly put.

It should be resisted and Mr Nott, to his credit, appears to be prepared to do just that. For the lessons of the Falklands war (which touch other modern governments than the British) run both ways. The task force to the Falklands could not have been mounted without ships, and large numbers of them, a large proportion of them fighting ships. That is one strike for the Navy lobby. But the recent conflict also showed that ships are vulnerable — part of the reason why successive British governments, whatever their first inclinations, have been cool about naval vessels as effective contributors to defence. (That they appear to burn so easily and well is nevertheless surprising.) Moreover, surface ships are likely for a long time to be more easily detected as floating metal objects on the surface of a foreign element, water, than they will be able to defend themselves with counter-measures. For the Navy lobby, that objection will not easily be turned.

0028-0836/82/250614-01$01.00

US looks to biological weapons

Military takes new interest in DNA devices

Washington

New evidence suggests that the US Army is planning a substantial expansion in its biological warfare research programme, and may be particularly interested in the potential role of recombinant DNA in the development of biological weapons.

Since signing the 1972 Biological Weapons Convention, the US government has maintained that all research on biological warfare is unclassified and strictly defensive. This research is carried out openly at the US Army Medical Research Institute for Infectious Diseases (USAMRIID) at Fort Detrick, Maryland.

The Office of Management and Budget (OMB) has confirmed, however, that the Army has included in its overall budget for medical research a sum for defensive biological weapons systems which is classified and which is totally separate from the published budget.

The budget office apparently became concerned that the amount the Army requested for this classified work was out of proportion to the Army's stated aims in this area. One biologist contacted by OMB in the course of its review of the Army request says that it involved "hundreds of millions of dollars". USAMRIID's budget is approximately $17 million.

Reports of the Army's interest in recombinant DNA appear to have originated from a request sent by the Army to the National Academy of Sciences several months ago. The academy has confirmed that the Army was sounding out its willingness to carry out studies on chemical and biological warfare that would involve classified materials. The academy's Assembly of Life Sciences decided not to participate in either classified work or any work involving biological warfare, but said it would consider doing unclassified studies on chemical weapons.

According to Professor Matthew Meselson of Harvard University, a long-time critic of the government's chemical and biological weapons policies, the topics the Army wanted the academy to study included the possible offensive uses of recombinant DNA technology in biological warfare, ostensibly for the purpose of better understanding how to defend against them.

The Army has since submitted a proposal to the academy for an unclassified study on the detection of and protection against mycotoxins, which the life sciences assembly apparently considers to be chemical agents, despite their biological origin.

The public affairs officer for the Army's medical institute, Norm Covert, said he was not aware of contacts with the academy nor of any US Army research on biological weapons apart from that conducted at Fort Detrick. The institute's programme involves only the development of medical knowledge about biological warfare agents, including detection, treatment and prevention. The only project that uses recombinant DNA is an effort to develop an anthrax vaccine by cloning in *Escherichia coli* the determinant of the protective antigen for anthrax.

Two years ago the Army advertised in *Nature* for proposals to clone the gene for human acetylcholinesterase. At about the same time, the Army received permission from the National Institutes of Health Recombinant DNA Advisory Committee to clone the determinant of a mild exotoxin from *Pseudomonas*.

Concern over what the Army may be planning has prompted two researchers to propose an amendment to the recombinant DNA guidelines that would specifically forbid "the construction of biological weapons by molecular cloning" (see *Nature* 17 June, p.527). It will be taken up at the next meeting of the Recombinant DNA Advisory Committee on 28 June.

In a statement from the Arms Control and Disarmament Agency which was cleared by the Department of Defense, the government told the committee that it has no objections to the amendment, but believes it to be unnecessary. The statement said that the Army's research programme "does not and will not involve research to create and screen 'new' organisms as potential biological warfare agents. Our research is, and will continue to be, limited to developing protective measures to recognized infectious agents which pose a biological warfare hazard." The statement also stresses that developing weapons for deterrence is not considered to qualify as one of the "prophylactic, protective or other peaceful purposes" for which research is allowed under the treaty. An official of the arms control agency said there was no evidence that the military was interested in going beyond defence research, and that the only classified material was information related to US vulnerability to biological attack.

Many of the same points were made in a

Changes for German cancer research

Heidelberg

The crucial meeting of the governing body (*Kuratorium*) of the German Cancer Research Centre seems to have passed off successfully on Monday (21 June). The centre will continue much on its present scale, but there will be substantial changes in the administration, thus vindicating the ambitions of Professor Hans Neurath, the late-director of the laboratory whose resignation last year precipitated the present crisis.

This week's meeting of the Kuratorium was called to consider the critical report of the independent commission of inquiry under Sir Michael Stoker that was published earlier this year and the response of the present staff, the new director and of the two governments (in Bonn and Stuttgart) that are involved.

The laboratory, strictly the *Deutsches Krebsforschungszentrum* (DKFZ), is West Germany's largest cancer research institute and has a staff of more than 1,000 deployed in 39 departments and a budget of nearly DM90 million (more than £20 million) a year.

The director, Professor Otto Westphal, and ministerial director Dr Fritz-Rudolf Güntsch presented the official response to the criticisms of the Stoker commission after the Kuratorium meeting. They agreed with the report's main contention, but said the commission had failed to appreciate the broad aims of the institute and also the administrative and legal constraints of West German institutions.

Within three months, Otto Westphal has succeeded in obtaining the cooperation of the Bonn ministry, the *Land* and the members of the institute in reaching a consensus on new measures. First, the role of scientists in the running of DKFZ is to be strengthened. A new scientific committee manned from outside will advise the Kuratorium on scientific projects, staffing, space problems and personnel. The Kuratorium will have eight scientific members out of 14 and, it is hoped, will be less concerned with administration and more with science. The executive committee of DKFZ must now either implement or rediscuss all decisions of the Kuratorium and not, as in the past, leave them in abeyance.

Second, deficiencies in reviewing procedures will be remedied, in particular by *ad hoc* commissions appointed by the Kuratorium and under the chairmanship of a Kuratorium member.

DKFZ is one of the few big cancer institutes without its own clinic. Westphal admitted the Cinderella role of clinical research in West Germany, which he put down to the structure of the medical institutions. He is optimistic about opportunities for unconventional extra-institutional collaboration. **Sarah Tooze**

0028-0836/82/250615-02$01.00

separate statement filed with the committee by Dr William Beisel, deputy for science at USAMRIID and the Defense Department's representative on the committee. An open question, however, is whether USAMRIID and the arms control agency are even aware of the Army's classified programme on defensive biological weapons. **Stephen Budiansky**

UK–Dutch agreement

Seeing stars

A far-reaching agreement for collaboration on major research projects was signed last week between the British Science and Engineering Research Council and its opposite number in the Netherlands, Nederlandse Organisatie voor Zuiver-Wetenschappelijk Onderzoek (ZWO). The immediate objective is to specify the rules under which the two research councils will collaborate on projects which involve expensive capital equipment.

A spokesman of the British council said last week that this development is a mark of the "steady convergence" of the policies of the two councils. Exactly a year ago, they signed an agreement on collaboration in astronomy under the terms of which ZWO will pay a fifth of the cost of the Las Palmas Observatory on Tenerife, receiving a fifth of the observing time in return.

Those administering the agreement say that the constructive benefits of the agreement are already apparent. Technical arguments by Dutch astronomers, for example, have led to the decision that the planned sub-millimetre telescope planned as part of the Las Palmas Observatory should be sited instead in Hawaii.

The intention now is that similar arrangements should be extended to other expensive projects, including the synchrotron radiation source at Daresbury, the British Starlink system for the common processing of astronomical data and possibly even the common use of major computer facilities. There is a possibility that the two councils may mount a joint project to build an improved neutron source for diffraction and other studies.

One administrative convenience in the new agreement is that it will not always be necessary for the costs of projects to be shared out one by one. Rather, when it suits the two councils, barter may replace the exchange of money.

The love affair between the two research councils has already led to the setting up of committees of which British and Dutch nationals are members. This has not yet, however, led to cross-membership of the principal policy-making committees, nor is there an immediate prospect of common grant-making procedures except where these are ancillary to some major project.

The Science and Engineering Research Council is the largest of the five in the United Kingdom, and exists to support scientific research at British universities and polytechnics. The terms of reference of ZWO are wider, extending to the humanities and social sciences. On the other hand, ZWO is not responsible for Dutch contributions to international projects (such as CERN), while the proposed Anglo-Dutch collaboration in the infrared satellite is similarly the direct responsibility of the Dutch government.

As a rule of thumb, the budget of ZWO in Dutch guilders is roughly equal to the budget of the Science and Engineering Research Council in pounds sterling (£1 = 4.7 guilders).

Cetus goes begging

Standard Oil Co. of California, an investor in Cetus Corporation, a leading California biotechnology firm, has elected not to fund the firm's plan to produce fructose commercially. Cetus hopes to find some other sponsor for the work, perhaps a sweetner manufacturer, instead.

The process would attempt to produce pure fructose at the same price as, or more cheaply than the main competitors — high fructose corn syrup (used in soft drinks) and sucrose. Although Cetus carries out research involving recombinant DNA techniques, the enzyme at the basis of the process was found using standard microbiological techniques.

Standard Oil's decision may represent a retrenchment by major oil companies in the biotechnology field generally. Investors in such firms are said to be more cautious now than they were one or two years ago, the budding recombinant DNA industry benefited from the enthusiasm — and dollars — of major firms. Standard Oil of California (Socal) owns 17 per cent of Cetus Stock. **Deborah Shapley**

Global systems analysis

Insult or injury?

The International Institute for Applied Systems Analysis (IIASA) in Vienna is going hawking for money. Abandoned recently by the US National Academy of Sciences, IIASA has suffered another blow. Its British member, the Royal Society, quit last week. Hopes for survival now centre on the institute's friends in the United States, who are trying to raise private money. Next month the director, Canadian Professor C. S. Holling, will visit Britain to attempt to rekindle interest.

For just over a decade, the institute's chief claim on public attention has been its unique status as the paradigm of the truly East–West research centre. Its collapse, now possible, could spell the end of an era or simply mean that the project was misconceived.

IIASA is being attacked on three fronts: political, financial and academic. Politically, the institute is a creature of detente, planned as a forum where East and West could tackle problems of global importance; but detente is dead, as is the Reagan government's interest in IIASA. And the £5 million budget (1982) is seen as an unnecessary luxury in hard times.

Academic criticism such as that which seems to have moulded the British decision has been more of a surprise. Sir Hermann Bondi, British member of the IIASA Council and also chairman of the UK National Committee for IIASA under the Royal Society, professes himself "horrified" at how few friends IIASA has on the committee. But Sir Hermann is by background a theoretical physicist, not a systems analyst, and must demur to professional opinion.

The UK committee seems to have been dismayed by the draft research plan for 1983, the first produced by Professor Holling. The committee had expected to see drastic pruning of the 24 projects current in 1981 under Soviet leadership of IIASA. Holling had reduced the number to nine, but these included the biggest of last year's projects, including for example the analysis of energy policies, the impact of industrial change, environmental regulation and institutions and regional and urban development. The committee decided that resources at IIASA were still being spread too thinly.

This opinion thus left the Royal Society unable to press the British government to continue membership. The Department of the Environment, the formal channel for the annual subscription, had already decided for internal reasons that it need no longer support IIASA. The Royal Society, Britain's formal member, without any obvious means of paying the subscription, resigned.

IIASA complains that the British committee does not understand the institute's objectives, which — as Holling has put it — are to provide policy-makers with "creative options". The British attitude is more pragmatic, IIASA staff members say, while there is very little overlap between the research interests of UK National Committee members and those of IIASA.

The institute is also offended that the British decision was taken when only the draft plan for 1983 was available. With the plan approved by the IIASA Council, the proposals can now be "fleshed out". Bondi has invited Holling to present his more detailed case personally to the UK National Committee next month. Holling will travel with Professor K. S. Parik of India and A. Wierzbicki of Poland to emphasize the international interest in IIASA work. It is too late for the Royal Society to reverse its decision, but Bondi hopes it is not too late to find another sponsor, however slim the chances.

In the past, IIASA has been represented by member institutions from 17 countries including the Soviet Union and the United

0028-0836/82/250616-02$01.00

States which had "Class A" membership and each paid 36 million Austrian schillings (£1.2 million). Lesser Class B members paid 5.4 million schillings. IIASA's nine projects for 1983 will now have to be found within a budget of 115 million schillings. IIASA's friends in the United States are looking for $1–1.5 million a year from 1984. **Robert Walgate**

Commercialization of research
So far. . .

Boston

Biological research progresses much more slowly than Wall Street transactions, Dr Walter Gilbert, professor of molecular biology at Harvard University reminded an audience of senior businessmen and brokers worth many millions in venture capital last week in the last lecture of his illustrious academic career. Gilbert, who in two weeks will take up full-time duties at the biotechnology firm Biogen, joined other eminent scientists in giving the executives a day-long, $300 per head tutorial on the rudiments of biotechnology. Several speakers addressed the question of how involved a university can get in the big business of biotechnology and still retain its integrity and autonomy.

Gilbert described the 30 years of basic research that prepared the ground for the recombinant DNA boom. Only in 1978, after three decades of work, could scientists attack the problem of making particular genetically engineered commercial products, and even the applied problems are proving very slow to solve. After the meeting, Gilbert reflected at length on the issues of the conference.

In the first place, Gilbert stated that aggressive university patent and "technology transfer" programmes are in fact superfluous. In the United States there exists no real technology transfer gap. "This is evident in the plethora of small biotechnology companies, and is due wholly to the effervescence of America's venture capitalists", he said, "who are the missing link in England."

Gilbert also took issue with the public perception that applied research is the intellectually weaker sister of basic research. And in fact, Gilbert continued, although everyone sees basic research as the wellspring of new scientific ideas, this is not exclusively true. "One reason universities encourage their faculty to consult for industry is that they realize that this is an important way for teachers to learn about the problems in the world."

With a foot in both industry and academia for five or more years now, Gilbert is in a unique position to comment on the threat to a laboratory of such cases as his own. Because "the student has absolutely no notion of why a professor puts him on a scientific problem", the threat of student exploitation is real. But Gilbert thinks that any commercial exploitation of a student will be exposed. In any case, the transgressor will be doomed, since commercially motivated work is often not academically meritorious and so he will not attract top-flight graduate students and government grants.

Dr Derek Bok, president of Harvard, gave the conference the clearest statement so far of the line universities will be taking on relations with biotechnology companies. Bok's "cleaner than thou line", as Gilbert later called it, was drawn up at conclaves of university presidents such as that last March in Pajaro Dunes (*Nature* 1 April, p 381). Universities are generally eager to aid in closing the putative "technology transfer gap", Bok said. He encouraged so-called "bilateral research agreements", such as the $50 million agreement between Hoechst and Massachusetts General Hospital. He underlined as justification for his position the current shortage of public research funding, the liberty inherent in a multiplicity of research sponsors and the fact that private support demands of the investigator little of the red tape that a federal grant does. Bok stated four provisos for such agreements: (1) that the sponsor cannot stipulate what the scientist studies, (2) that bilateral agreements must be published, (3) that a firm must guarantee a discoverer's right to publish his findings, and (4) that a firm must get preferential patent rights only when it has clearly funded the work involved.

Gilbert later called the patent issue the messiest in all the debates. Patents on discoveries made by university workers should go to the discoverer himself, not to a private supporter and particularly not to the university. "The superficial rationale behind Harvard's current position of retaining patent rights on its employees' inventions is that this is in the public interest", Gilbert said, "but actually the university has only one motive: remuneration." In Gilbert's view, the university should renounce patent rights and any other procedures which would lead to its motives appearing suspect. The disinterestedness that would result is one of the university's strongest assets. In the long run, renouncing patents would actually make it richer, since over the centuries such a climate of disinterestedness is what has made universities worth endowing.

Bok overtly proscribed investment by the university in a firm in which one of its faculty members had a stake. He recounted the example of the firm in which Dr Mark Ptashne is involved, in which Harvard once considered a joint venture. Although it "takes an imaginative soul to dream up a gift a university will refuse", Bok mused, this was an offer which Harvard had to turn down to preserve good public relations, faculty morale and general academic values. Dr Ptashne later retorted that in fact Bok had got it wrong — that it was the university and not Ptashne's company that made the first overture.

Cases in which a faculty member has a stake of money or time in a biotechnology firm are the hardest to legislate. The concern of Harvard, Bok said, is with "what occupies a professor's mind when he wakes up in the morning". Since this information is inaccessible, there must be certain guidelines. For instance, at Harvard and at Massachusetts Institute of Technology no professor may be a corporate executive and retain his faculty post. Nor may he hold a "significant number of shares" in a job-related company. ("The main flaw in Bok's position," Gilbert later said, "is in his argument on shareholding. What constitutes a 'significant number of shares'? There is only one clean line and that is no shares at all.") Bok said that on the other hand the university permits its faculty members to spend up to 50 days per year away from their laboratories consulting with industry (at a going rate of $200–$2,000 per day). He warned that it is as important that a university should be seen to avoid the dangers in faculty or university relations with biotechnology firms as that it should actually avoid them.

James Aisenberg

British biotechnology
Imperial poised

Imperial College London has found an ingenious way of dragging itself by its bootstraps into the brave new world of biotechnology. A few weeks after the college arranged that its fermentation pilot plant should be transferred to a private company financed by TDC Developments Limited and called Imperial Biotechnology Limited, a venture capital firm, the college is in the market for three staff members — a professor (whose stipend will be provided by the Leverhulme Trust) and two lecturers, whose salaries will be met by the fees earned from contracts with outside bodies based on the use of the pilot plant.

The chairman of the Centre for Biotechnology that will result, Professor Brian Hartley, says that the development at Imperial College shows that universities can still embark on novel undertakings when general funds for universities are restricted. In these days, he says, a college that has three vacant posts on offer in such a field is in a unique position. He is undismayed that the local branch of the academics' union (the Association of University Teachers) has complained that new academic posts should not be created when the holders of other academic posts are being threatened with redundancy.

The Imperial College pilot plant, built somewhat before its time was ripe at the instance of the late Sir Ernst Chain and until quite recently considered something of a white elephant, and has turned out to be a marketable asset. Imperial College has, however, reserved 20 per cent of the time available at the pilot plant in lieu of shares that would otherwise have been available.

0028-0836/82/250617-02$01.00

Exploiting genetics

Washington

Congress keeps muttering about legislation to regulate links between universities and industry on the exploitation of genetic engineering Last week, Congressman Albert Gore's oversight committee held two days of hearings on the conflicts of interest that have arisen or may arise The occasion gave university representatives an opportunity to parade local solutions they believe will prevent a recurrence of recent controversies

In several well-publicized cases, university researchers pursued work in outside corporations that closely paralleled their academic research, and the financial arrangements between the parties, suggested conflicts of interest (In the recent Calgene case, for example, a plant biologist at the University of California at Davis received a grant from Allied Chemical The researcher was also vice-president of Calgene, a biotechnology firm in which Allied subsequently purchased a 20 per cent interest (see *Nature* 4 March, p 6)

Recent large grants by corporations to Massachusetts General Hospital and Washington University, St Louis, have heightened concern over possible conflicts with traditional freedom of academic inquiry

Roderick Park, vice-chancellor of the University of California at Berkeley, testified — as did representatives of the University of Wisconsin, Stanford University and the University of California at Davis — that the universities are themselves concerned Draft principles under consideration at Berkeley, for example, would prohibit any research on campus "whose benefits to education and research are small", would require sponsors of research to allow free publication of all results, and would "scrutinize", but not necessarily forbid any arrangement that involved sponsorship of campus research by a company in which the researcher held a financial interest

At the hearings, critics of university–industry ties had a chance to point a finger at some of the more notorious conflict-of-interest cases and called for federal guidelines on financial disclosure Albert Meyerhoff of the Natural Resources Defense Council (NRDC) testified that the universities have already proved incapable of confronting these issues themselves The Pajaro Dunes conference failed to accomplish its purpose of drawing up ground rules for industry-sponsored research, he said Meyerhoff also criticized the Stanford faculty for having rejected a proposal requiring disclosure of conflicts of interest

NRDC called for federal legislation that would require universities receiving federal research funds to adopt financial disclosure rules for faculty Researchers should disclose any financial interests they have in research sponsors and also any interests in companies that could benefit from their research, NRDC said

No such legislation has been introduced But Representative Albert Gore (Democrat, Tennessee), who conducted the hearings, has been increasingly concerned over the effects of new industry–university ties

Stephen Budiansky

Professor Hartley says that the new research and teaching centre will have two chief lines of enquiry — the engineering of microorganisms able to digest wood and wood-like natural materials (including sugar-cane and baggasse) into usable chemicals, and the development of enzyme electrodes or sensors by which means the activity of a biochemical enzyme can be coupled directly to a semiconductor device "We may be able to make microprocessors that can smell", Professor Hartley said

On his own position at the centre, Professor Hartley says that there is no conflict between his chairmanship of a new academic centre and his membership of the scientific advisory board of Biogen, the Swiss-based company He explains that his contract with Biogen allows him to keep confidential his work within Imperial College and vice versa He considers that the centre, now that it is a going concern, will be able to apply successfully for a subvention from the fund earmarked by the University Grants Committee for the support of biotechnology in British universities (see *Nature* 20 May, p 173)

Research council visitors

No-gag gag

Allegations that foreign scientists visiting British research council laboratories are subject to serious constraints on their freedom of expression seem to be a storm in a teacup What is at issue is whether the terms on which visiting scientists agree to work in British laboratories muzzle public criticism of council policy

Mr Stanley Alderson, who describes himself as a writer and human rights campaigner, claims that foreign scientists visiting the Agricultural Research Council (ARC), Medical Research Council (MRC), Science and Engineering Research Council (SERC) and Social Sciences Research Council (SSRC) are asked to sign an agreement, known as Form Y, accepting the conditions of work undertaken by British employees of the councils As well as agreeing to observe safety arrangements, patent regulations and conditions governing publication of research work, visiting scientists are also required "during (their) visit and afterwards not to mention the Council's name in any public controversy" Mr Alderson says that the last condition is inspired by Section 2 of Britain's Official Secrets Act and contravenes the human rights guidelines of several international organizations

The issue may be taken further Not only does Mr Alderson plan to write to the national newspapers but Lord Avebury has written to Sir James Gowans, Secretary of MRC, drawing his attention to Form Y Lord Avebury professed himself astonished that "any self-respecting scientist would give such an undertaking which on the face of it puts a gag on visiting scientists" and prevents them from ever commenting publicly on the policy of the research councils

The research councils are clearly surprised and bewildered to find themselves at the centre of such a maelstrom "Form Y? Notes for the guidance of visiting scientific workers? What is it? We've never heard of it " Form Y is in fact used only by MRC, visiting scientists are asked to sign it in exchange for access to MRC facilities

An MRC spokesman said that Form Y had nothing to do with the Official Secrets Act and merely expressed work conventions understood by employees of any establishment The phrase which requires that MRC should not be mentioned in any public controversy conforms with the guidance given to its own employees and was inserted to avoid embarrassement — when making a public statement the individual should make it clear that he is speaking on his own behalf and not stating council policy unless he has obtained official approval in advance MRC employees, he claimed, are not prevented from expressing their opinions as private individuals

Visiting scientists at ARC and the National Environment Research Council (NERC) are asked to sign a form covering standard conditions — patent rights, health regulations, publication conditions — but there is no clause referring to the councils in any public controversy NERC employees and overseas visitors who come for two or three months do have to sign Section 2 of the 1911 Official Secrets Act and Section I(2) of the 1920 Act SERC has no official form at all, and the only restriction on visitors is that covering patent agreements

What seems clear is that the obligation, explicit or understood, on an employee not to speak publicly on behalf of his employing body without official approval does not limit his right to speak out on any issue whatsoever as long as he makes it clear he is expressing a private opinion What is not so clear, however, is whether an individual, employed by or visiting a research council, can publicly criticize the policy of the council without the fear of being disciplined

Jane Wynn

0028-0836/82/250618 01$01 00

Plans for research

Europe awakes

Brussels

"It is of course the duty of the European Commission to be ambitious", commented one British official cynically on the European Commission's new plans to increase EEC science cooperation Next week's council meeting of the research ministers (on 30 June) will be sitting in judgement on a strategy for the future of European science policy, the fruit of two years of conferences and discussions that followed the 30 May mandate of 1980, when Mrs Margaret Thatcher won Britain's refunds from EEC's budget and when the other nine member states agreed to strengthen and broaden EEC's activities in the industrial, social and scientific fields

Step by step, the European Commission has been subtly building up momentum by calling unusually frequent research council meetings during the past nine months Previously, research ministers met every two years but, pushed by the European Commissioner with the science portfolio, Vicomte Etienne Davignon, the EEC ministers met in November and March and will see each other next week and again in the autumn

The momentum for change is being created not so much by decisions as by ministerial debates intended to set the mould for a recasting of the priorities and design of EEC research programmes The meetings so far have established the important role of research and development in achieving industrial competitiveness with the Japanese and Americans in new technologies and the economic and social dangers of failing to do so

Davignon wants a new research and development programme with five objectives improving the competitiveness of agriculture and industry, the management of raw materials and energy resources, the quality of life in terms of health and safety and the environment, helping the less developed countries, and stimulated science and technology in the Community

EEC spending — 645 million European currency units (£1,150 million) in the 1983 budget — represents only 2 7 per cent of the budget as a whole Sixty-nine per cent of the total budget goes on energy, 13 5 per cent on industrial competitiveness, 9 per cent on the quality of life, 2 per cent on agriculture and 1 per cent on raw materials But research and development employs more staff (3,000) than any other directorate general in the Commission if the staffs of the Joint Research Centres are included

The new strategy would be based on five massive long-term programmes One of these, thermonuclear fusion, is already in operation The others would tackle information technology, biotechnology, the stimulation of industrial and agricultural competitiveness and research intended to aid developing countries

The broad outlines of the information technology programme, dubbed "Esprit", have already been sketched out (see *Nature*, 3 June, p 352), work on the other programmes may begin after the forthcoming council meeting The reason for this is significant Instead of laboriously budgeted and detailed programmes being presented to infrequent ministerial meetings, the planning would in future work from the top downwards

There would be a new body, the Committee for the European Development of Science and Technology (CODEST) bringing together twenty of the great and the good in the world of science, technology and industry to select research objectives It would interpret a master plan agreed by ministers, who would specify in advance the resources available A plan for encouraging the research judged necessary would then be devised by CODEST and the Commission, on which basis proposals from research teams capable of achieving the objectives in question would be chosen But independent experts drawn largely from industry would actually run the programmes The first steps towards a long-term programme would be pilot projects funded by EEC

In this way, with CODEST operating as a think tank, the Commission as its administrative arm and the day to day running handled independently, the political rows that sully programmes such as the reactor safety project Super-Sara would be avoided There would no longer be haggling and long delays over the budgets of programmes that had already been worked out in detail

The realization of the Commission's plans depends on the momentum established by the forthcoming council meeting It is suggested in Brussels that Davignon will continue merely to be humoured because in a time of bugetary restraint EEC research, whatever other purposes it may fulfill, is not on its past record the best value for money Few deny that there is a drastic rethinking of the rationale behind European research and development

Jasper Becker

Soviet mathematics

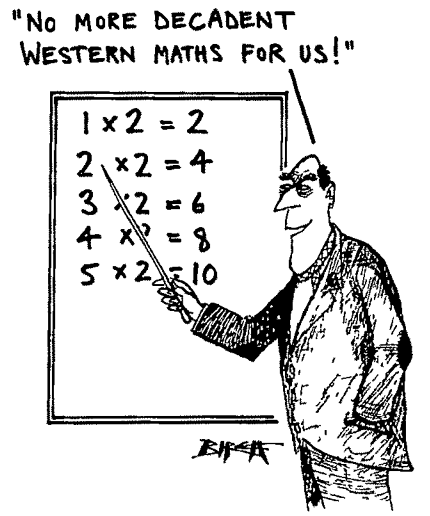
No more sets

The Soviet Union is to introduce a new mathematics syllabus in the senior classes of secondary schools The objective according to Dr M I Kondakov, president of the Academy of Pedagogic Sciences, is to resolve the major debate which has been raging among Soviet educationalists since the introduction of the present syllabus into Soviet schools a few years ago

The new syllabus was part of a major overhaul of school education, begun in the mid-1960s Special emphasis was placed on an adequate grasp of general principles and on pupils' ability to reason for themselves Learning by rote was to be replaced by an increased amount of independent work In particular, mathematics was to be firmly based on set theory

Initially, the new syllabus appeared to be a success In the first three school years the proportion of poor achievers in mathematics dropped from 15 per cent in 1967 to 3 per cent in 1975 But in the middle and senior years, the situation was less happy In 1977, when the first batch of students to have experienced seven years of the new syllabus took their school-leaving examinations, it was found that a majority had failed to grasp the new concepts

One particular failing was a marked lack of spatial imagination, attributable either to the introduction at too early an age of a purely academic approach or else to the fact that, so far from working "independently", the pupils were issued with ready-drawn diagrams to save time in class The



curriculum seems also to have nullified the avowed purpose of the reforms by using a different terminology from that used in science classes, so that pupils found themselves unable to solve quadratic or trigonometric equations arising in physics classes

These examination results triggered a major debate in the Soviet educational press, with the president of the Academy of Sciences, Dr Anatolii D Aleksandrov, and the prestigious mathematician L I Pontryagin criticizing the syllabus Their equally prestigious colleagues S L Sobolev and L V Kantorovich, however, cautiously supported it

The debate raged for two years (a detailed analysis of the arguments appears in the current issue of *Journal of Curriculum Studies*), and, perhaps significantly, teachers' complaints centred around difficulties in implementing the syllabus (lack of classroom aids and teachers' handbooks) rather than the actual content, although their main difficulties seem to have arisen from the highly theoretical orientation of the syllabus

Parallel with this debate, the weekly *Literaturnaya Gazeta* launched its own

0028-0836/82/250619-02$01 00

discussion, which ran for some months and in general proved to be critical of the current emphasis on mathematics as such One correspondent depicted a future with a computer on every classroom desk and suggested that with the advent of the computer as "co-respondent", the long "marriage" of mathematics and science was ripe for divorce

Professor M Evgrafov, of the Mathematics Institute of the Academy of Sciences went further "Not a single result" in pure mathematics, he claimed, had had any practical application for the past fifty years, and the current insistence on a mathematical training for scientists was "wanton", "rigid" and "useless" In the future, he said, problem-solving will be increasingly left to programmers whom he described as "persons of middle-level qualifications"

In 1980, Pontryagin shifted the whole debate from the educational and cultural journals to *Kommunist*, the leading Party theoretical monthly His criticisms thus took on the force of tacit party policy, and have undoubtedly played a part in shaping the revised syllabus, which will give greater emphasis to practical applications and seems designed to train the "middle level" technicians of whom Evgrafov spoke

The signs are that the new syllabus will downgrade set theory (except, optionally, in geometry), eliminate "difficult" terminology and symbolism and play down theoretical precision in favour of an "intuitive and descriptive" approach with an emphasis on new mathematical technology, including computers **Vera Rich**

Falkland minerals

An agreement of sorts between Argentina and the United Kingdom on the title to submarine mineral rights off the Falkland Islands appears to have been reached on the eve of the occupation of the islands, and to have been embodied at the end of April in a resolution of the United Nations conference on the Law of the Sea The resolution adopted by the conference (but with Argentina's dissent) says that where there are disputes about the sovereignty of territories (such as the Falkland Islands), mineral resources shall be exploited as if the wishes of the inhabitants were paramount

This resolution appears to be a compromise reached after the rejection by Western powers, France and the United Kingdom in particular, of a draft article of the treaty that would have given each disputing party an effective veto on mineral exploitation, and which used the weaker form of words "for the benefit of the inhabitants"

The rejected article of the treaty, described as "anti-colonialist", would also have given the United Nations an important role to play in the regulation of disputed marine mineral rights The resolution eventually adopted by the conference, which will not have the force of international law, plainly suits the British book much better

British Antarctic Survey

A good wind

On the principle that "it's an ill wind ", the British Antarctic Survey is hoping that the war over the Falkland Islands will bring back good times again Meanwhile, the survey is congratulating itself that those of its scientists left stranded on South Georgia after the occupation by Argentine forces have made good use of their extra six weeks on the island

The supply vessel that would normally have picked up the scientific parties from South Georgia in April, the RRS *Bransfield,* also seems to have done useful work The ship was apparently in the Chilean port of Puerto Arenas at the occupation of the Falkland Islands, collecting spares for diesel engines and other gear In the succeeding weeks, it was able to monitor radio transmissions between Port Stanley, on the Falkland Islands, and the mainland of Argentina

During the ship's spell in Antarctic waters (below 60 degrees south), the ship served as a relay station between the British bases in Antarctica, and even kept in touch with those isolated on South Georgia by means of relayed signals from the base at Signey one the South Orkney Islands Normally, communications with the British Antarctic bases are relayed to Cambridge, England, through the radio and telex station at Port Stanley, out of action since early April

For most of the Antarctic programme, the interruption caused by the Falklands war should be small, as most of the bases were resupplied just before the occupation of the Falklands On South Georgia, however, where there has been a weather station in continous operation since 1944 and more recently a summer biological station on Bird Island nearby, a season's work is likely to be lost

Apart from the difficulties of operating its two ancient ships in hostile waters, most of the British Antarctic Survey's problems are monetary The numbers of people on its books have been shrinking for the past decade, and now amount to about 300, of whom 120 are scientists The budget for the current year is about £6 million, a modest decline compared with the previous year without allowing for inflation Even so, it has been possible to rebuild two of the three bases on the Antarctic Peninsula, while there is an ambitious plan to rebuild the base at Halley, on the Weddell Sea

What the survey now hopes is that the British government's renewed interest in the South Atlantic, and the practical need to find peaceful activities that will occupy British nationals there, will prompt more generous budgets

Chinese census

First head-count begins next week

Experts applaud revised plan

Washington

The only modern census of China's estimated 1,000 million people, who make up one-quarter to one-fifth of the world's population, will take place on 1 July China's last census was in 1964, but Western demographers consider its results suspect The earlier count in 1953 is similarly in doubt So the July count is a demographic event of major historical proportions

At the 1953 census, Mao Tse-Tung officially governed 583,603,417 people The number was growing rapidly in accordance with Chinese family customs and Mao's dictum of 1949 that "revolution plus production can solve the problem of feeding the population Of all things in the world, people are the most precious " But by 1957, the net increase was 20 people per thousand per year and the official line changed Mao said the population should be stabilized, perhaps at 600 million people

That figure had probably already been exceeded and subsequent famines, political upheavals and other changing conditions have increased uncertainty as to how many people China has Age, sex and regional breakdowns are critical for its centralized state planning But an accurate count, besides helping the government, will among other things show whether the government's past claims for success in population control are true

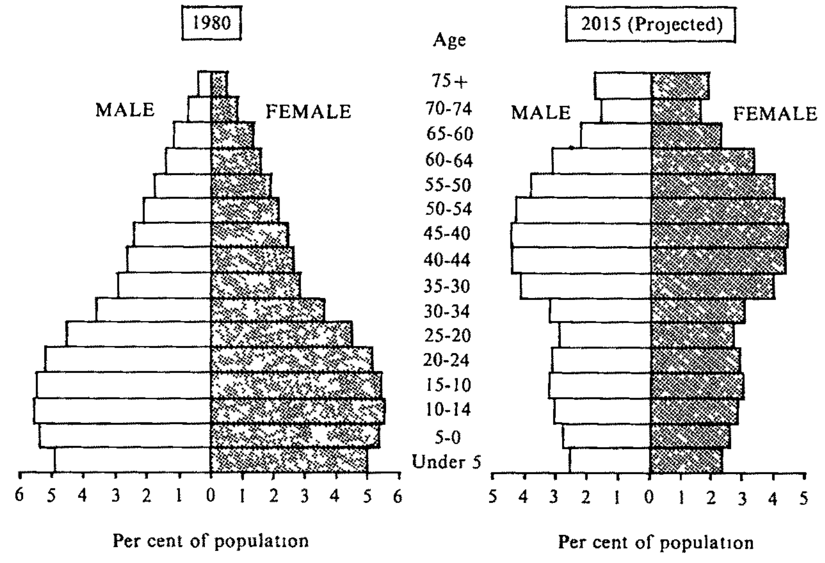
But the scale of the enterprise is staggering How do you count 1,000 million people? Previous censuses in China were carried out literally with hand tabulation and abacuses For the 1982 census, the government has acquired 20 IBM 4331 E Series computers and one IBM 4341 computer with funds from the United Nations Fund for Population Activities (UNPA), and has bought eight Wang computers on its own

The five million or so enumerators will be office workers, staff of local institutions, rural accountants and record keepers In a test count in the county of Wuxi, enumerators were required to have at least a junior middle school education The census forms will have 19 categories to fill in, some by code, some by ordinary writing (The 1964 census form had eight categories, the 1953 census had five) Representatives of each household will be summoned to registration points to tell the enumerators how many people were in the household at midnight, 1 July Or they will be reached by mail, or visited on their

0028-0836/82/250620 02$01 00

boats Post-enumeration checks will be made, followed by hand tabulations, after which the information will be sent into the computer for the region Eventually the information will be centralized in Beijing The schedule requires enumeration in early July and a provisional count by October 1982 of total population, numbers of males and females, population by province and other features A machine tabulation of 10 per cent of the national sample will be ready in the spring of 1983, the government says, while the remaining tabulation could be completed and released in 1984 The Chinese are taking a pragmatic, self-critical approach to the census This reflects not only the temper of the current regime but the tastes of the census's chief architect, Li Chengrui, head of the State Statistical Bureau (SSB) He and his colleagues have been collecting advice and assistance from outside China for some years now They have changed their plans to provide the most modern count possible and say they will make the results fully available

Present and future population (both estimated)



Source *Intercom Vol 9, p 13 (Population Reference Bureau, August 1980)*

Credence

If the preparations for the count on 1 July are breathless, China was even less well prepared in 1979 for the census it then planned for 1980 Western demographers were surprised at the credence placed by the Chinese in their earlier counts, in 1953 and 1964, and critical of the plan that the same techniques should be used again They were especially sceptical of the claimed error rate of 0 0014 per cent

What seems to have been learned in the past few years of international consultation is that the Chinese government needs to know accurately how many people it has to govern, but that there are limits to the accuracy of any census Their project may be helped by the character of Chinese society Chinese neighbourhoods are close, while party officials know who lives in this or that house

The cities will be the hardest problem Many city dwellers have moved there illegally, and may remain hidden from the counters There is also a risk that families that have broken family planning rules may conceal the fact, while local officials knowing that too many babies have been born in their districts may conspire to keep young children off the forms

Even if such problems of bias can be overcome, the 1 July census will, for the time being, stand on its own The earlier counts in 1953 and 1964 are known to have been inaccurate, and so inaccurate that they cannot be extrapolated forwards to 1982 Thus it will not be possible to assess the effectiveness of the family planning campaigns begun in 1956, especially because the Great Leap Forward and the succeeding Cultural Revolution made a hash of many state statistics One difficulty during that period was the doctrine that statistics are "cold" and "too professional" and the accompanying injunction that statisticians should submit their work to the scrutiny of their superiors and should "be humble" in accepting corrections to it

Part of the interest in the new census and in the age structure of the Chinese population it will provide is the light it may throw on events such as the widespread famine that in 1960 eliminated countless numbers of the population, and which may be apparent in the size of the age group that will be 22 this year But the sheer size of the Chinese population will be the chief interest Sample counts suggest that it is 1,000 million, but some US experts think that an underestimate by about 200 million

Publicity

This year's census is being conducted very differently from that in 1964, whose existence leaked out only when a British cancer study cited figures from the previously unknown census This time, the project is being undertaken in what amounts, for China, to a glare of publicity

A major point of interest in the 1982 census is what it will tell about the success of recent government population control measures The crude birth rate stood at 37 births per thousand until Mao's first efforts at population control, it dropped to 20 per thousand in 1960, and rebounded up to 43 6 per thousand in 1963 The more serious controls of the 1970s have brought this rate down in 1978 the Chinese claimed the crude birthrate was 18 34 per thousand So, when the results are published in October, Western experts will study them closely to see how many people turn up in the 0 to 4 and 5 to 9 age group The figure reflects the importance of population control measures to China It represents how the Chinese population might change by the year 2015 if each woman in her life had an average of 1 5 babies The comparison with China's traditional age profiles, on the left, show how important demographics can be for state political and economic planning

Resistance

Indeed, if China turns out to have ten million or one hundred million people more than previously thought, the leadership will be tempted to use stronger measures to enforce population control China has been importing food but has little hard currency to buy a lot of it from abroad So the implication of extra people is ominous Paradoxically, the successful population control measures of recent years have just been undercut by the state's effort to increase food production By allowing farmers to keep their produce and turn over a pre-set quota to the state, local authorities have lost a major source of control over households, including the number of children a couple has At the same time the farmers, given a chance to keep what they produce, have an incentive to have more children for additional labour So resistance to population control, which is embedded in the Chinese culture, may stiffen just when more babies are the last thing the country needs

Deborah Shapley

China percentage age distribution 1953–80

Ages (years)	1953	1964	1980 estimate
0–4	15 59	14 52	9 11
5–9	10 94	13 65	12 55
10–14	9 39	12 52	12 93
15–19	9 10	9 01	9 05
20–24	8 24	7 37	8 29
25–29	7 76	7 31	9 06
30–34	6 88	6 77	7 19
35–39	6 41	5 97	5 35
40–44	5 56	5 17	5 02
45–49	5 01	4 47	4 59
50–54	4 24	3 84	4 11
55–59	3 61	3 27	3 59
60–64	2 90	2 56	3 08
65–69	2 03	1 69	2 47
70–74	1 37	1 07	1 80
75–79	0 61	0 55	1 84
80+	0 35	0 26	—

Sources 1953, Chinese census 1964, Chinese census 1980, US Census Bureau estimates based on official Data Model B, May 1980

CORRESPONDENCE

Nuclear hazard

SIR — While it is true, as N Bielsten states (*Nature* 25 March, p 284), that thousands of people have died over the past 20 years in dam failures, it is naive to compare this to the record of the nuclear power industry The fact remains that nuclear power reactors represent the only form of civilian technology with the potential to kill several million people in a single accident

While one can argue *ad infinitum* about the probability of such an event, we must admit that there have been several near misses and, as more reactors are built and as present reactors age and become more prone to failure, the likelihood of such an accident is increasing For the nuclear power industry to congratulate itself for the fact that no such accident has yet occurred is reminiscent of the fellow who jumped off the Empire State Building and who was heard to exclaim as he passed each floor "So far, so good"

ROBERT J YAES
Memorial Hospital for Cancer and Allied Diseases,
Sloan-Kettering Institute for Cancer Research,
Cornell University, New York, USA

Fear not frenzy

SIR — Your alliterative headline "Freeze frenzy hits US" in *Nature* of 29 April (p 790) was misleading, and gave a wholly wrong impression about feelings here There is no frenzy we are just frightened, and rationally so I don't want nuclear bombing to destroy me and my family, my home, our laboratories and libraries and universities and so on, but if any major nuclear war starts there seems very little doubt that all this would happen in the first few hours I imagine that reasonable people would feel similarly in Britain, the Soviet Union and elsewhere At a recent meeting in here in San Diego, some 3,000 people gathered to discuss nuclear disarmament The discussions were sober — and sobering — but not frenetic

RALPH A LEWIN
Scripps Institution of Oceanography,
University of California, San Diego,
California, USA

Tangible results

SIR — Martin Raff (*Nature* 25 February, p 642) has observed a disquieting trend towards restricted distribution among scientific colleagues of biological materials

When research results are in the form of biological material, such as desired gene sequences cloned in a plasmid vector, simple publication of such research results may not be adequate, or simply not most efficient, to advance science

Such research results should be made available openly and promptly to scientific colleagues, whether in academe or industry This is the basic premise of the Tangible Research Property policy recently issued at Stanford to affirm appropriate scientific practice (*Nature* 25 March, p 283) The policy specifically adjures against withholding distribution for commercial reasons Cell lines, including hybridomas, are distributed to colleagues Broad distribution of Tangible Research Property (TRP) can be done in a fashion which reasonably protects commercial rights The optimal mode of protecting TRP would involve strictly limited distribution or secrecy, not options for a university to follow

With the foregoing as background, let me turn to Dr Raff's concern about open distribution of hybridomas and monoclonal antibodies There is a great demand for monoclonal antibodies, as is well known to readers of *Nature* Indeed, this issue of *Nature* may have several advertisements offering various monoclonal antibodies for sale Producing and distributing monoclonal antibodies to large numbers of users is beyond the capability or funding of the normal university laboratory Private companies provide a significant service to science by making available well characterized monoclonal antibodies of consistent quality to academic and industrial scientists for their research As monoclonal antibodies find their way to clinical applications, the role of industry will be of even greater significance

At the point of clinical application, some measure of proprietary protection for commercial sales is needed by industry to enable its considerable investment to bring therapeutic products to market Exchange of biological materials or other products should and can be accomplished in a fashion that preserves such proprietary protection and thus avoids jeopardizing the public's access to therapeutic or other products derived from such biological materials This may involve prior (to distribution) agreement by receiving scientists to follow steps not to destroy those proprietary rights needed for commercial development

Molecular biology is in transit to an era of increased practical applications for public benefit The commercialization aspect of this can be depressing, as Dr Raff observes, but the beneficial aspect should not be overlooked

NIELS J REIMERS
Director, Technology Licensing,
Office of Technology Licensing,
Stanford University,
California, USA

Billions and upwards

SIR — All support to your "ban the billion" campaign — what use is a scientific word that can mean 10^9 or 10^{12}? Since the ambiguous terms billion, trillion, quadrillion, quintillion, sextillion, septillion, octillion, nonillion (see 6th edition of *Concise Oxford Dictionary*) all end in -llion (and this is therefore a suitable suffix for all large number words) could we not have general agreement on gigallion 10^9, terallion 10^{12}, petallion 10^{15}, exallion 10^{18}? As far as I am aware, none of these words has been used before, and certainly not to denote different numbers from those suggested here

Larger numbers can be constructed by using two (or more) of the internationally recognized prefixes, for example, megaexallion for 10^{24}, exaexallion for 10^{36} Cubi- will take us even further, cubiexallion $(10^{18})^3$ being 10^{54}

DAVE E PARRY
Bell College of Technology,
Hamilton, Scotland, UK

Boldly going

SIR — Wallis (*Nature* 15 April, p 598) appears to support Hoyle's view that there has been insufficient time for the development of life systems on the Earth In particular, he says that 2,000 proteins of a specified character are required to catalyse the reactions involved, and that these would not have appeared by random assortment of amino acids into polypeptide chains in the time available

This argument would be valid if the development of life on planets like the Earth depended upon the simultaneous presence of about 2,000 different polypeptide chains, each containing a uniquely specified group of about 6 amino acids to catalyse a particular biological reaction This is not true

The emergence of life did not have to await the appearance of a particular collection of proteins It emerged from the random set of proteins that were already present on the primitive Earth

Earth's biochemistry is a by-product of the continuous irradiation of the primitive ocean, producing chains of reactions and processing organic molecular systems back down to the free energy gradient As has been shown experimentally, irradiated systems of appropriate composition contain all the organic forms (proteins, nucleotides, carbohydrates and so on) required for Earth's biochemistry The huge variety of peptide chain segments present would provide catalytic centres for many of the reactions involved The self-replicatory and evolutionary properties of systems of proteins and nucleotides then ensures that the peptide-catalysed chains of reactions play an increasing role in the flow down the free-energy gradient, perhaps by the mechanisms described by Black[1] At some arbitrarily chosen stage in this process of development, we should recognize the system as a form of life

Wallis's argument does, however, show that no two planets in the Universe can have the same biochemistry Captain Kirk and the crew of the *Enterprise* are no more likely to find life forms with the Earth's biochemistry than to find them speaking English

A E ROUT
British National Oil Corporation,
London W1, UK

1 Black, S *On the Thermodynamics of Evolution* (*Perspectives in Biology and Medicine*, Vol 21(3), Spring 1978)

PhD job centres

SIR — In his letter published in *Nature* on 13 May (p 98), A F W Coulson describes an information exchange for PhD students in the life sciences and speculates why this has not been done previously in other fields May I point out that in physics such a compendium was first published by the Institute of Physics in 1969 at the request of the Standing Conference of Professors of Physics The Sixth Edition has recently appeared

I too, Sir, have often wondered why similar volumes have not appeared for other subjects

L COHEN
Institute of Physics,
London SW1, UK

NEWS AND VIEWS

Gravitational wave backgrounds and the early Universe

from B J Carr

COULD the detection of a background of gravitational waves tell us what the Universe was like during its first few moments? This is the intriguing question raised in a recent paper by Adams, Hellings, Zimmerman, Farhoosh, Levine and Zeldich in the *Astrophysical Journal* (**253**, 1-18, 1982) Those interested in detecting gravitational waves have usually focused attention either on bursts of radiation (such as might be generated by supernova explosions or black hole formation) or on continuous radiation (such as might be generated by binary stars and pulsars) Strenuous efforts to detect these types of radiation, using resonant bars and laser interferometers, are already afoot Attention is now turning to the possibility of detecting a background of gravitational radiation, analagous to that of electromagnetic radiation Such a background would have to have been generated at a cosmological redshift — that is, well before the present epoch, and perhaps before galaxies formed — so any success in detecting it could give vital information about the early Universe

One way in which a background of cosmological gravitational waves could arise is from the superposition of many individual bursts generated astrophysically at some time in the past If the waves were generated early enough (necessarily at redshifts exceeding 1) and if the events producing them were sufficiently numerous, then the gravitational bursts should overlap in the sense that their duration t should exceed their average separation Δt This would produce a stochastic background with an effective dimensionless amplitude larger than that of the individual bursts by a factor $(t/\Delta t)^{1/2}$ Overlapping bursts could have been generated by various postgalactic events They could also have been produced by pregalactic events, for example, by the formation of a population of pregalactic black holes Such holes might today provide the 'missing mass' known to reside in clusters of galaxies and galactic haloes The period and amplitude of the resulting background will, in this case, have been uniquely determined by the holes' mass and density formation epoch, so its detection could help us ascertain the nature of the missing mass

Another possibility, and one that is particularly relevant to the Adams *et al* paper, is that the background could be primordial in the sense that it goes all the way back to the big bang and is part of the initial conditions of the Universe For example, it is possible that quantum processes in the early Universe produced a thermal background of gravitons with a present temperature of the order of 1K, analagous to the thermal background of 3K microwave photons which was discovered in 1965 Unfortunately, it seems unlikely that such a thermal background of gravitons could ever be observed A more interesting possibility is that primordial waves might have resulted from purely classical processes if the early Universe, instead of having the smooth structure observed today, was chaotic In this case, the waves would not look like radiation at early times because they would have a wavelength larger than the Universe's particle horizon size (the distance travelled by light since the big bang) Also, their dimensionless amplitude, instead of being small, would be of the order of unity and thus severely distort the background space-time This contrasts to the situation with an astrophysically generated background, where the waves could never have a significant cosmological effect

There is no doubt that solutions to Einstein's equations which contain incipient gravitational waves in this sense ought to exist, although the complexity of the equations makes the identification of such solutions non-trivial However, Adams *et al* have succeeded in providing a formalism with which to describe plane gravitational waves in Bianchi cosmological models (The Bianchi models are homogeneous in the sense that they look the same everywhere, but anisotropic in the sense that they look different in different directions) By confining attention to plane-wave solutions, which break the homogeneity of the Bianchi model only in the direction of wave propagation (the z-axis), they manage to find exact solutions which do indeed exhibit the sort of properties anticipated above at early times the Universe is highly inhomogeneous, with the anisotropy parameter depending on the z-coordinate, while at late times one gets waves travelling in the z-direction They thus show explicitly how the chaotic behaviour near the initial singularity can be transformed into gravitational waves Admittedly, the chaotic early Universe is unlikely to possess the plane symmetry assumed by Adams *et al* A genuine stochastic background, with overlapping waves coming from all directions, would need to derive from a much more complicated initial behaviour Nevertheless, it is an important first step and the study of more complicated models will doubtless follow

An interesting characteristic of gravitational wave backgrounds, whether primordial in origin or generated astrophysically, is that they could have considerably longer periods than the kinds of waves generated by sources in the present epoch The period of primordial waves is more or less arbitrary but could, in principle, be anything up to the age of the Universe, pregalactic black hole events could generate waves with present periods up to a year Although such long-period waves could not be detected with conventional bar-type detectors, they could be using more exotic methods For example, they might be detected by Doppler tracking interplanetary spacecraft a spacecraft would be jiggled by gravitational radiation and this would be registered as a change in the frequency of the radio waves which are routinely used to track it Long-period waves could also be detected by their perturbations on planetary orbits Since the precision with which planetary orbits and spacecraft motions can be monitored is already close to that required to detect possible gravitational wave backgrounds, it is conceivable that gravitational wave astronomy will start in space rather than on Earth

If waves of primordial origin were discovered, it would obviously be important to ask what informationa they could yield about the initial state of the Universe If the early Universe really were as simple as the Adams *et al* model presumes, that is, inhomogeneous in only one direction, then measurements made by only three spatially separated observers would suffice to infer the initial state However, the Universe, if chaotic at all, is unlikely to have been that simple and so in practice many more observers would be required to draw any cosmological conclusions Nevertheless it is intriguing that, in principle, monitoring the positions of planets and satellites could answer questions about the initial singularity It should be noted that in the Adams *et al* solutions, the gravitational wave amplitude decreases with time and thus becomes increasingly difficult to detect Unfortunately, this decrease only occurs on a cosmological time scale, so this feature may not convince organizations which fund gravitational wave research of the urgency of such measurements! □

B J Carr is at the Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA

0028-0836/82/250623 01$01 00

Stop-go proteins

from Graham Warren

SOME ten years ago the discovery of the signal sequence[1,2] showed how, in principle, the cell could 'tag' proteins destined for secretion When a secretory protein is synthesized the first part of the growing polypeptide chain to emerge from the ribosomal complex is a 'signal' or 'leader' sequence that directs the ribosomal complex to the membrane of the endoplasmic reticulum (ER) There the signal sequence is removed and the rest of the protein synthesized as it is transferred across the membrane into the cisternal space

Such discoveries were made possible by the development of cell-free systems that mimicked the *in vitro* transfer process Signal sequences have subsequently been found on all nascent proteins initially inserted into the ER membranes What has remained a mystery, however, is the way in which the signal sequence carries out its function Other cellular proteins must be involved but their study requires a painstaking dissection and reconstitution of the ER membrane Progress in this area has until recently been slow, but now two ER proteins have been characterized[4,5] one stops synthesis of the protein until a suitable site for transfer across the ER membrane has been found and one, described in this issue of *Nature* (see p 647), 'docks' the protein at a suitable site

It has been known for several years that microsomal vesicles (derived from dog pancreatic ER) lose the capacity to transfer secretory proteins *in vitro* when extracted with high salt solutions Activity can be restored by a component of the salt extract — a protein complex containing six different polypeptide chains[9] — but the precise function of the protein complex was unknown Peter Walter and Günter Blobel have recently shown[4,5] that it binds strongly to those polysomes synthesizing secretory proteins and stops further translation, but binds only weakly to free ribosomes and has no effect at all on polysomes synthesizing cytoplasmic proteins Translation of secretory proteins is stopped after the nascent chain has reached a length of about seventy amino acids Forty of the amino acids will be buried within the large ribosomal complex and thirty will have emerged from the ribosomal complex — the same length as the signal sequence for the secretory protein under study The signal sequence is rich in leucine which can be replaced in the cell-free system by a more hydrophilic analogue, β-hydroxyleucine When this is done the protein complex can no longer effectively inhibit synthesis of the now-modified secretory protein The authors therefore conclude that the complex recognizes the signal sequence and have termed it the signal recognition protein (SRP) Though much work needs to be done to determine the individual contribution of the six polypeptides to the inhibition of translation, the complex has been found to perform the same function with other secretory proteins[6] (and see p 647) and at least one membrane protein[9]

The block in translation can be released by addition of the salt-extracted microsomal vesicles and subsequent synthesis is then coupled to transfer The ER proteins responsible for releasing this block were unknown but one protein essential for *in vitro* transfer[10] had been previously characterized by Meyer and Dobberstein It is released from salt-extracted microsomal vesicles by protease treatment and has a molecular weight of 60,000 Antibodies to this proteolytic fragment[11] showed it to be derived from an ER membrane protein with a molecular weight of 72,000 that is entirely restricted to the ER *in vivo* The work reported in this issue of *Nature* (p 647) now shows that the purified proteolytic fragment specifically releases the translation block induced by SRP The membrane bound form of the protein has the same effect This means that the block on synthesis produced by SRP will not normally be released until the SRP has bound to the membrane protein The protein thus acts to 'dock' the ribosomal complex at a suitable site on the ER

To accord with the overall scheme for the homing step proposed by Walter and Blobel[4,5] at least some of the SRP should be free in the cytoplasm ready to bind to polysomes synthesizing proteins initially destined for the ER SRP was, however, characterized from a salt-extract of microsomal vesicles The cytoplasmic form has now been detected (see p 647) using a sensitive assay and significant amounts of SRP have been found in the cytoplasm of reticulocytes and dog pancreatic cells

The homing step explains how the ribosomal complex becomes specifically bound to the ER membrane The signal sequence does not play a direct role in this step but acts through SRP This may have an additional advantage since the hydrophobic signal sequence would not be free to interact with any cellular membranes other than the ER Once it is free to interact with the lipid bilayer, the homing step will have ensured that this bilayer will be that of the ER The inhibition of translation is clearly an important part of SRP's function, to

Graham Warren is in the European Molecular Biology Laboratory, Heidelberg 6900

SPEAKING on 27 May at a Royal Society discussion meeting on Cell Membranes and Glycoprotein Synthesis, Dr Gunter Blobel of the Rockefeller University gave advance notice of a new and unexpected twist in the long running saga of the signal sequence It now seems that the signal recognition protein, whose role is described in the accompanying article, has a specific ribonucleic acid associatied with it

That the ribonucleic acid plays some, as yet undefined, part in the process of ensuring that secretory proteins are secreted is suggested by the inhibitory action of ribonuclease on the ability of the signal recognition protein to arrest the translation of secretory proteins (see accompanying article) The sedimentation coefficient and partial sequence of the ribonucleic acid associated with the signal recognition protein make it all-but certainly identical to 7S RNA, the sequence of which has just been described by two independent laboratories (E Ullu, S Murphy & M Melli *Cell* **29**, 195, 1982 and W Li, R Reddy, D Henning, P Epstein & H Busch *J biol Chem* in the press) Intriguingly, 7S RNA is closely related to the '*Alu*' family of RNAs, the genes for which exist in many thousands of copies scattered throughout the mammalian genome *Alu* RNAs, although present in abundance in most cells, are of unknown function It could be that the role of one subset has now been unearthed

At the same meeting Blobel acknowledged the demise of a different twist in signal sequences His classic signal sequence is at the amino terminus of a secreted protein and is cleaved off as the protein traverses the membrane Later it emerged that ovalbumin, although a secreted protein that competed with others for insertion into microsomes *in vitro*, did not subsequently lose an amino-terminal peptide In 1979 Blobel's laboratory published a paper (V R Lingappa, J R Lingappa & G Blobel *Nature* **281**, 117) claiming that the signal peptide of ovalbumin was internal, and suggesting a location around amino acid 250 of the 385-residue chain Now, as Blobel acknowledged, it looks as if that result was in error One is instead to believe the conclusion of William Braell and Harvey Lodish (*J biol Chem* **257**, 4578, 1982) that the signal sequence of ovalbumin is on the amino-terminal side of residue 150 and thus not exceptional

But, according to an earlier paper of Braell and Lodish (*Cell* **28**, 233, 1982), what is not true for ovalbumin is true for the erythrocyte anion transport protein which really does have an internal signal sequence somewhere around its middle

Peter Newmark

0028-0836/82/250624-02$01 00

stop translation must be a complicated process reflected in the polypeptide complexity of SRP

The advantage of stopping translation until a site on the ER has been found is not hard to see Without this block translation would continue and the ribosomal complex would only have a few seconds to reach a suitable site for transfer of the protein If it did not reach a suitable site in time the protein would be completed and discharged into the cell cytoplasm This could be very damaging particularly if, for example, these proteins were degradative enzymes By stopping translation SRP make sure that there is sufficient time to find a suitable site

These arguments may not apply in prokaryotes which are much smaller (so that less time would be needed to reach a membrane) and have only one membrane (the plasma membrane) through which proteins are initially transferred It is even not clear that proteins must be transferred during synthesis[13] and in one case (M13 phage coat protein) transfer can occur after synthesis and apparently requires only a suitable lipid bilayer and signal peptidase[14,15] It would therefore be important to see whether prokaryotes have the equivalent of SRP and a docking protein since their presence would argue strongly in favour of transfer during synthesis

The homing step is distinct from, and precedes, the transfer of the nascent protein across the ER bilayer The ribophorins and the signal peptidase[3] may be involved in this step but no other cellular proteins have been characterized Reconstitution techniques now need to be applied to the transfer step It will be more difficult to study because the ER lipid bilayer will have to be broken down and built up again Nevertheless, the striking success of reconstitution techniques in revealing the functions of signal recognition protein and docking protein make the attempt well worthwhile □

1 Milstein, C , Brownlee, G G , Harrison, T M & Mathews, M B *Nature new Biol* 239, 117 (1972)
2 Blobel, G & Dobberstein, B *J Cell Biol* 67, 852 (1975)
3 Sabatini, D D , Kreibich, G , Morimoto, T & Adesnik, M *J Cell Biol* 92, 1 (1982)
4 Walter, P , Ibrahimi, I & Blobel, G *J Cell Biol* 91, 545 (1982)
5 Walter, P & Blobel, G *J Cell Biol* 91, 551 i 557 (1982)
6 Stoffel, W , Blobel, G & Walter, P *Eur J Biochem* 120, 519 (1981)
7 Warren, G & Dobberstein, B *Nature* 272, 569 (1978)
8 Walter, P & Blobel, G *Proc natn Acad Sci U S A* 77, 7112 (1980)
9 Anderson, D J , Walter, P & Blobel, G J *Cell Biol* 93, 501 (1982)
10 Meyer, D I & Dobberstein, B *J Cell Biol* 87, 498 & 503 (1980)
11 Meyer, D I , Louvard, D & Dobberstein, B *J Cell Biol* 92, 579 (1982)
12 Rothman, J E & Lodish, H F *Nature* 269, 775 (1977)
13 Josefsson, L -G & Randall, L L *Cell* 25, 151 (1981)
14 Silver, P , Watts, C & Wickner, W *Cell* 25, 341 (1981)
15 Watts, C , Silver, P & Wickner, W *Cell* 25, 347 (1981)

Related genes can have unrelated introns

from Athel Cornish-Bowden

IN eukaryotic gene families, such as those that code for the globins, the number and position of the introns has been found to be conserved across each family and, presumably, during evolution A very different picture emerges from new work reported in this issue of *Nature* (p 655) by Woo and his collaborators They have studied the introns in the genes for two distantly related proteins, human α_1-antitrypsin and chicken ovalbumin, and have found that there is no correspondence between the three introns of the ovalbumin gene and the seven of the α_1-antitrypsin gene This result calls into question the idea that the arrangement of introns provides a record of the original structure of a eukaryotic gene, instead it suggests that introns may have been introduced into genes since the divergence of the vertebrates

Ovalbumin is the principal protein in chicken egg white, whereas α_1-antitrypsin is a human plasma protein involved in the control of elastase, individuals deficient in it have a high risk of lung disease There was no reason to expect any sequence similarity between the two proteins and the observation that they are in fact 24 per cent identical was thus a surprise The degree of similarity is far too high to be ascribed to chance, as Hunt and Dayhoff (*Biochem biophys Res Commun* 95, 864, 1980) estimated that such a possibility would occur with a probability around 10^{-41}

If chance is excluded, there are two ways in which the similarity between ovalbumin and α_1-antitrypsin can be explained, either the two proteins have diverged from a common ancestral protein, estimated by Hunt and Dayhoff to have existed about 500 million years ago, or they have converged to similar sequences as a consequence of selection In the former case, the lack of correspondence between the arrangements of introns in the two genes needs to be explained, in the latter it does not because no correspondence would be expected Woo and collaborators say that it is impossible to determine conclusively whether the two proteins arose by convergent or divergent evoluion I believe them to be unduly cautious in this regard and that the possibility of convergent evolution can be discounted

Although examples of convergent evolution of protein structures do exist, they are observed at the level of catalytic mechanism and not in amino acid sequence In the best established case, that of the bacterial proteinase subtilisin and the mammalian serine proteinases such as chymotrypsin, the catalytic mechanisms are essentially the same but there is no sequence similarity beyond a use of the same kinds of residue in the catalytic action By contrast, if the similarity between ovalbumin and α_1-antitrypsin arose by convergent evolution it would imply convergence to a high degree of similarity over the whole lengths of two proteins without even, apparently, a similar physiological function If this were substantiated its shattering effect on current ideas about protein structure and evolution could hardly be exaggerated, it would be comparable with the effect of the first reports of introns in eukaryotic genes and overlapping genes in viruses on our understanding of gene structure

The chicken ovalbumin gene contains seven introns, all of them located in the 5′ half of the mRNA, whereas the three introns in the human α_1-antitrypsin gene are all in the 3′ half of the mRNA Only one intron in the ovalbumin gene occurs in even approximately the same position as any intron in the α_1-antitrypsin gene the introns in question do not show significant sequence similarity either with one another or with the corresponding coding regions of the other genes It seems inescapable that a substantial reorganization of the intron structure has occurred since the time of the common ancestral gene shared by the ovalbumin and α_1-antitrypsin genes

Could the differences in intron structure be a consequence of random losses of introns during evolution from an ancestral gene containing at least ten introns? If there were exactly ten, of which a random three were deleted on the lineage leading to the ovalbumin gene, and a random seven were deleted on the lineage leading to the α_1-antitrypsin, the probability that the two descendent genes would have no introns in common would be $3!7!/10!$, or 1/120 This probability is hardly small enough to rule out random deletion of introns as an explanation, but it is small enough to suggest that alternative explanations should be considered, not only in this case but also in that of the actin gene family, for which a much smaller degree of intron variability has been observed Woo and collaborators suggest that some at least of the introns may have been introduced into the ovalbumin and α_1-antitrypsin genes after they diverged in evolution If so, the arrangement of introns must be a more

Athel Cornish-Bowden is in the Department of Biochemistry, University of Birmingham, PO Box 363, Birmingham B15 2TT

0028-0836/82/250625-02$01 00

recent aspect of gene structure than has been supposed Light will perhaps be shed on this question by investigation of the intron structure of the gene for human antithrombin III, another plasma proteinase inhibitor in the super-family that contains ovalbumin and α_1-antitrypsin □

Membrane kissing

from A W Cuthbert

CELLS that respond very rapidly to chemical signals such as neurotransmitters change their ionic permeability by altering the functional state of ion channels already present in the membrane An alternative but improbable way of changing permeability might be to insert patent ion channels into the membrane from a preformed store within the cell However for cells that respond only slowly to signals, as with some hormones, such an alternative way of changing permeability could have advantages For instance, a sustained response might follow only a transient exposure to hormone with return to the prestimulated condition being governed by membrane turnover processes

Lewis and de Moura report in this issue of *Nature* (p 685) what may be a convenient way to study the insertion of ion channels into cellular membranes from an intracellular source The tissue they have exploited is rabbit urinary bladder epithelium, which transports sodium electrogenically from the apical to the basolateral side They "punched" stripped preparations of this tissue by subjecting the apical side to a series of rapid changes in hydrostatic pressure As a result, transepithelial resistance was reduced and short-circuit current increased Punching caused more than a tenfold increase in current together with a significant increase in the sodium selectivity of the apical face The current, like the normal sodium transport, was sensitive to the pyrazine diuretic amiloride The authors suggest that punching causes cytoplasmic vesicles containing sodium channels to collide and kiss the apical membrane, resulting in the transfer of sodium channels to the apical surface

The urinary bladder, of course, undergoes cyclical changes in volume as it collects urine Lewis and de Moura also show that when the bladder is distended mechanically the membrane capacitance remains rather constant while the mass per unit area is decreased This, they argue, is due to the insertion of new membrane material from an intracellular source keeping the capacitance constant at around 1 μF cm^{-2} Similarly, when the cells are allowed to swell in a hypotonic medium there is an increase in apical membrane area which results in a 74 per cent increase in apical membrane capacitance The effects both of punching and of exposure to hypotonic solutions were blocked by cytochalasin B, perhaps indicating the involvement of microfilaments

An unexplained finding was that although punching increased apical sodium permeability there was no increase in capacitance If vesicles rich in sodium channels are incorporated into the apical membrane when it is punched, then equivalent amounts of membrane with fewer sodium channels must be removed on rebound

Undoubtedly these experiments will give investigators another way to manipulate the apical membrane of transporting epithelia As with many discoveries, their importance is not necessarily that they are new but rather that they are timely Insertion of channels into epithelial membranes as a way to increase permeability has been proposed for both sodium and water in recent years As far back as 1968, Nutbourne[1] applied small hydrostatic gradients to the apical surface of frog skin and recorded marked increases in sodium transport, yet this attracted little attention although the technique used was very similar to that described by Lewis and de Moura It is interesting to re-read the discussion of that paper written at a time when the mechanisms for sodium ion translocation across the apical face were poorly understood

The mammalian urinary bladder, like numerous other tight sodium-transporting epithelia, responds to aldosterone with an increase in ion transport Early studies[2,3] and more recent electrophysiological measurements[4] have shown that the sodium permeability of the apical surface of epithelia increases under the influence of this mineralocorticoid Current fluctuation analysis in potassium-depolarized toad bladders has shown that the unit conductance and average lifetime of individual sodium channels are not changed but that there is an increase in the number of electrically active sodium channels[5] Although the action of aldosterone requires DNA-dependent RNA synthesis and protein synthesis it does not follow that aldosterone actually causes the synthesis of new ion channels The metabolism of the cell may be altered in such a way that preformed ion channels stored in cytoplasmic vesicles are inserted into the apical membrane, without ruling out an additional effect on synthesis Recruitment of electrically silent channels seems also to be responsible for the change in the sodium permeability of the apical face in amphibian urinary bladders treated with antidiuretic hormone (ADH)[6], but whether this occurs from electrically silent channels in the apical membrane or by fusion of cytoplasmic vesicles is not known True, there is an increase in the membrane capacitance in toad bladder in response to ADH[7] which might be expected from the insertion of new membrane However, there are two problems with this interpretation First, Lewis and de Moura did not see an increase in capacitance in their punched tissues although sodium transport was markedly increased Second, ADH also causes an increase in water permeability of the toad urinary bladder which may be responsible for the capacitance change

Painstaking morphological studies using freeze-etching have shown that ADH causes the appearance of particle aggregates in the cytoplasmic face of the apical membrane[8,9] There is strong correlative evidence that the particle aggregates are associated with increased water permeability, but whether the aggregates are formed by translational movement in the membrane or by insertion from the cytoplasm is not clear, although the increase in capacitance favours the latter view Interestingly, methohexital blocks the water permeability response to ADH without affecting the sodium transport response[10] Whether the whole of the capacitance increase is also blocked by methohexital remains to be explored

Membrane macromolecules must be made within the cell so it is not unlikely that

A W Cuthbert is Sheild Professor of Pharmacology in the Department of Pharmacology, University of Cambridge, Hills Road, Cambridge CB2 2QD

1 Nutbourne, D M *J Physiol Lond* **195**, 1 (1968)
2 Crabbe, J *J clin Invest* **40**, 2103 (1961)
3 Sharp, G W G & Leaf, A *Physiol Rev* **46**, 593 (1966)
4 Nagel, W & Crabbe, J *Pflügers Arch ges Physiol* **385**, 181 (1980)
5 Palmer, L G , Li, J H -Y , Lindemann, B & Edelman, I S *J Membrane Biol* **64**, 91 (1981)
6 Li, J H -Y , Palmer, L G , Edelman, I S & Lindemann, B *J Membrane Biol* **64**, 77 (1981)
7 Stetson, D L , Lewis, S A & Wade, J B *Biophys J* **33**, 43a (1981)
8 Kachadorian, W A , Levine, S D , Wade, J B , Di Scala, V A & Hays, R M *J clin Invest* **59**, 576 (1977)
9 Chevalier, J , Bourguet, J & Hugon, J S *Cell Tissue Res* **152**, 129 (1974)
10 Levine, S D , Levine, R D , Worthington, R E & Hays, R M *J clin Invest* **58**, 980 (1976)
11 Fambrough, D M *Physiol Rev* **59**, 165 (1979)
12 Posner, B I , Josefsberg, Z & Bergerson, J M *J biol Chem* **254**, 12494 (1978)
13 Cezard, J P , Conklin, K A , Das, B C & Gray, G M *J biol Chem* **254**, 8969 (1979)
14 Maze, M & Gray, G M *Biochemistry* **19**, 2351 (1980)
15 Schilling, E E , Goldenberg, H , Morré, D J & Crane, F L *Biochim biophys Acta* **555**, 504 (1979)
16 Doyle, D D , Wong, M , Tanaka, J & Barr, L *Science* **215**, 1117 (1982)
17 Aceves, J , Cuthbert, A W & Edwardson, J M *J Physiol, Lond* **295**, 477 (1979)
18 Van Driessche, W & Lindemann, B *Nature* **282**, 519 (1979)
19 Cuthbert, A W & Shum, W K *Proc R Soc* **B189**, 543 (1975)
20 Cuthbert, A W , Edwardson, J M , Bindslev, N & Shadhauge, E *Pflügers Arch ges Physiol* **392**, 347 (1982)
21 Eldrup, E , Mollgard, K & Bindslev, N *Biochim biophys Acta* **596**, 152 (1980)

0028-0836/82/250626-02$01 00

molecules *en route* for the membrane, and in increasingly complete stages of assembly, will be found in the endoplasmic reticulum, Golgi and cytoplasmic vesicles Reservoirs of a number of molecules destined for the cell membrane and showing modified immunoreactivity, enzymatic potential or binding properties have been identified intracellularly[11-15] A recent example, relevant to this discussion, is the identification of a saxitoxin-binding protein in frog heart cytosol as a putative sodium channel[16] Amiloride analogues have been used to identify components in sodium-transporting epithelia with properties consistent with those of apical sodium channels Binding measurements in intact tissues, such as frog skin[17], have given site densities (130 μm^{-2}) comparable with fluctuation analysis[18] (50 μm^{-2}) However, in isolated cells[19] and homogenates[20] where the ligands have access to intracellular sites, much greater site densities have been obtained Finally, in the epithelium lining the avian coprodaeum, rod-shaped particles, tentatively identified as sodium channels, have been seen in freeze-fractured membranes, both apical and intracellular[21]

Thus there is a body of very different sorts of evidence which suggests that transporting epithelia have a reservoir of vesicles containing sodium channels poised for insertion into the apical membrane Whether these vesicles can 'kiss and make up' the increase in sodium permeability of the apical surface in response to punching may be very hard to prove □

Inhumation: how ants and other invertebrates help seeds

from Andrew J Beattie and David C Culver

WHAT happens to seeds once they come to rest following dispersal? Seeds are scattered widely in the environment by wind, water, animals and explosive mechanisms and yet the sites where successful germination and growth can occur are often patchy, scarce and limiting[1,2] Some possess mechanisms such as hygroscopic bristles for self-burial[3], while others are effectively buried by falling into soil crevices or where wind- or water-borne soil and debris accumulate[2] Many seeds do not require burial for germination, but a common fate for those remaining on an exposed surface is to be eaten[4] Are there mechanisms that increase the probability with which they will end up in a better site?

Recent work on ant dispersal of seeds (myrmecochory) strongly suggests that a major selective advantage is provided by the relocation of seeds to ant nests which are richer than surrounding soils in essential nutrients such as phosphorus, potassium and nitrogen[5-7] Seeds end up on the soil or litter surface following dispersal and are then removed by ants which are attracted to nutritious tissues (elaiosomes) on the seed coat Once in the nest, the elaiosomes are removed and the seeds abandoned, intact and viable, either in an old gallery or in the refuse pile[8]

In a series of experiments with two myrmecochorous species of *Viola*, the fates of seeds taken into nests by ants were compared with the fates of seeds planted at random in the same habitats and it was found that seedling emergence is nearly three times more likely from nests Furthermore, the seedlings emerging from ants nests are larger, and the probability of their surviving the first two years is considerably higher (0 14 against 0 03)[7,9]

Additional ant-mediated effects appear in a study of another myrmecochore, *Sanguinaria canadensis* This species can persist in disturbed habitats, where the primary ant dispersers are absent, by means of vegetative propagation but, in undisturbed forests, where seed-dispersing ants are present, profound changes in adult density, dispersion and patterns of reproduction are brought about[10] Further beneficial effects also come from the greater chance that seeds relocated to nests will escape the attention of predators[4,11]

Myrmecochory may be much more widespread than has been thought European and North American temperate deciduous forests and Australian sclerophyll shrublands have often been regarded as its lone strongholds[12,13], but we now know that it occurs commonly in a very wide variety of plant community types worldwide, from low to high latitudes and elevations in both the Northern and Southern Hemispheres[9] Myrmecochorous species are known from 80 plant families (including grasses and cacti) and may be trees, shrubs, vines, herbaceous perennials, annuals, hemiparasites, parasites and epiphytes Myrmecochores constitute 35 per cent of species in certain habitats, and up to 76 per cent of emergent stems[13-15].

Invertebrate groups other than ants may have similar effects on the fate of seeds Scarab beetles (*Scarabeidae*) in the Cerrado of Brazil collaborate to roll and bury fruits The larvae mature by eating the pulp and the viable, intact seeds remain below ground[16] Ground beetles (*Carabidae*), common in many habitats[17,18], store seeds in the lining of their tunnels The function of this behaviour is obscure, but the seeds appear to remain unharmed[19-21]

Earthworm burrows and casts are known to be richer in critical plant nutrients than surrounding soils[5,22] and many kinds of seeds are passed intact and viable embedded into the cast[23] Snails too may be of importance A series of experiments in which seeds were placed with captive snails showed that the overwhelming majority pass through the digestive tract to be deposited in the faeces intact

We suggest 'inhumation' (Latin· *humare* = to cover with earth) may provide a useful term for the process described above While this process has been described for seeds passing through the guts of vertebrates such as birds, bats and various other mammals[24-27], the inclusion of major and abundant invertebrate groups, such as ants, beetles, earthworms and snails, may make the process of inhumation far more important than previously supposed Mortality in plant populations is generally greatest among seedlings deprived of suitable microsites or sufficient seed reserves Inhumation may circumvent these dangers to some degree by increasing the probability that seeds are relocated to microsites already provisioned with nutrients □

Andrew J Beattie and David C Culver are in the Department of Ecology and Evolutionary Biology, Northwestern University, Evanston, Illinois 60201

1 Solbrig, O T, Jain, S, Johnson, G B & Raven, P H *Topics in Plant Population Biology* (Columbia University Press, New York, 1979)
2 Harper, J L *The Population Biology of Plants* (Academic, New York, 1977)
3 van der Pijl, L *Principles of Dispersal in Higher Plants* (Springer, New York, 1969)
4 Heithaus, E R *Ecology* **62**, 136-145 (1981)
5 Petal, J in *Production Ecology of Ants and Termites* (ed Brian, M B) 293-325 (Cambridge University Press, New York, 1978)
6 Davidson, D W & Morton, S R *Oecologia* **50**, 357-366 (1981)
7 Culver, D C & Beattie, A J *Am J Bot* **60**, 710-714 (1980)
8 Culver, D C & Beattie, A J *J Ecol* **66**, 53-72 (1978)
9 Beattie, A J *Abh Verh naturw Ver Hamburg* (in the press)
10 Pudlo, R J, Beattie, A J & Culver, D C *Oecologia* **146**, 32-37 (1980)
11 O'Dowd, D J & Hay, M E *Ecology* **61**, 531-540 (1980)
12 Serander, R *K svenska Vetensk-Akad Handl* **41**, 1-410 (1906)
13 Berg, R Y *Aust J Bot* **23**, 475-508 (1975)
14 Beattie, A J & Culver, D C *Ecology* **62**, 107-115 (1981)
15 Handel, S N, Fisch, S B & Schatz, G E *Bull Torrey bot Club* **108**, 430-437 (1981)
16 Gottsberger, G K *Abh Verh naturw Ver Hamburg* (in the press)
17 Schaller, F *Soil Animals* (University of Michigan Press, Ann Arbor, 1968)
18 Borror, D J, DeLong, D M & Triplehorn, C A *An Introduction to the Study of Insects*, 4th edn (Holt, Rinehart & Winston, New York, 1976)
19 Alcock, J *Bull ecol Soc Am* **54**, 45-46 (1973)
20 Kirk, V M *Ann ent Soc Am* **65**, 1426-1428 (1972), **66**, 513-518 (1973)
21 Manley, G V *Ann ent Soc Am* **64**, 1474-1475 (1971)
22 Edwards, C A & Lofty, J R *J appl Ecol* **17**, 533-543 (1980)
23 McRill, M & Sagar, G R *Nature* **243**, 482 (1973)
24 Ridley, H N *The Dispersal of Plants Throughout the World* (Reeve, Ashford, 1930)
25 McKey, D *The Coevolution of Animals and Plants* (eds Gilbert, L E & Raven, P H) 159-191 (University of Texas Press, Austin, 1975)
26 Lieberman, D, Hall, J B, Swaine, M D & Lieberman, M *Ecology* **60**, 65-75 (1979)
27 Janzen, D H *Ecology* **62**, 593-601 (1981)

0028-0836/82/250627 01$01 00

A beginner's guide to major histocompatibility complex function

from Polly Matzinger and Rose Zamoyska

THE glycoproteins encoded by the major histocompatibility complex (MHC) are guides by which cells of the immune system distinguish both each other from all other tissues, and also all cells of the body from non-cells (such as bacteria, viruses, toxins, etc.). These two different modes of discrimination require two different sets of MHC-coded cell surface molecules, Class I and Class II.

Class I molecules are characteristic of cells

An example of the importance of the recognition of cells versus non-cellular material can be seen in the immune response to a virus infection, which is fought mainly by a subset of T lymphocytes called T killer cells. This set of lethally destructive lymphocytes can kill virus infected cells before synthesis of new viruses is complete, thus preventing the spread of the infection. But how do T killer cells distinguish between *virus infected cells,* which they can destroy, and free virus particles, which they are powerless to stop?

The T cell population consists of millions of constantly circulating lymphocytes, each one bearing receptors able to bind to one particular three-dimensional structure. This results in a heterogeneous population which can recognize a wide variety of foreign molecules. Normally these T cells are quiescent and must be activated to become effective killer cells. Their activation is not a simple process but requires the participation of a special accessory cell called the "antigen-presenting cell". A virus entering the body is picked up by the presenting cells which process it and present it to the roving T cell population. Those few T cells which have receptors specific for viral structure will bind to the presenting cell. Although there may be resting T cells which can bind to virus alone, these are not activated. Only those T cells which recognize *both* the Class I cell surface molecules and the viral structures on the presenting cell are induced to proliferate and differentiate into active killer cells. Since induced T cells maintain their specificity, and since only those T cells which bind both viruses and Class I cell surface molecules are induced, the resulting activated killer cells are committed to the recognition of cell-bound virus. Furthermore, activated killer cells are extremely precise in their recognition of both virus and MHC molecules, discriminating between closely related strains of virus and also distinguishing between different Class I alleles. An animal bearing a particular Class I allele 'a', for example, will produce killer cells which lyse infected target cells expressing 'a' but will ignore infected target cells bearing 'b' alleles. This obligate recognition of foreign molecules in the context of MHC proteins is called 'MHC restriction' of T-cell recognition.

Nearly all tissues express Class I molecules and, if infected, can be destroyed by active T killer cells. However, in many species, Class I molecules are not expressed by mature red blood cells (having no DNA, red cells cannot support virus replication) nor by sperm (although they are foreign invaders, sperm are usually welcome).

Class II molecules are characteristic of cells of the immune system

While T killer cells cope with cell-bound virus, antibodies secreted by activated 'B' cells deal with free virus particles. In order to be activated, resting B cells require signals from a subset of T cells called (not surprisingly) T helper cells. If a T helper cell specific for virus is to help a B cell make antibody against the virus, it must have the capacity to recognize not only the virus but also the B cell. It would be futile for T helper cells to send signals to cells outside the immune system (a virally infected liver cell, for example) so, unlike T killer cells, T helper cells do not recognize virus in the context of the ubiquitously expressed Class I molecules. Instead they are limited to the recognition of virus in the context of Class II molecules, which are expressed mainly by B cells and antigen-presenting cells. The activation of T helper cells may be described in a similar way to the MHC-restricted activation of T killer cells. Circulating T helper cells encounter virus structures on the surface of a presenting cell. They are activated as a result of interacting with virus and Class II molecules on the presenting surfaces, and will subsequently assist virus-specific B cells carrying the same Class II molecules.

Class II genes were originally discovered as immune response (IR) genes, so called because animals which carried different MHC alleles differed in their ability to make immune responses to certain foreign molecules. This now appears to be explained by the 'guidance' function of the MHC molecules. It is known that a foreign molecule may be recognized more efficiently with one Class II allele than with another. Animals bearing class II alleles that are poorly recognized in association with the foreign molecule will not generate activated T helper cells and will consequently be unable to produce strong antibody responses. It has been suggested that an MHC molecule might be inefficient at presenting a particular antigen because the two molecules do not form a good complex on the surface of the presenting cell or, alternatively, that a complex does form, but for various reasons there are no helper T cells able to recognize it.

How do MHC molecules activate T cells?

There are at least two possible ways in which MHC proteins could act as T-cell activators.

1. Like hormones, they might work directly, activating any T cell which binds them.
2. The binding of MHC molecules by T cells may trigger the antigen-presenting cell to send activation signals to the bound T cell. T cells binding virus alone would therefore not trigger the antigen-presenting cell and would not receive the activation signal. The fact that both Class I and Class II molecules are transmembrane proteins, and the observation that not every cell expressing these molecules can act as T-cell inducers, support the latter possibility. Further evidence that MHC molecules act as presenting cell triggers comes from transplantation experiments.

MHC genes, being highly polymorphic, are the major barriers to tissue transplantation. T cells distinguish between the products of self and foreign MHC alleles with great precision, destroying any tissue bearing foreign MHC products. The magnitude of this reaction is greatly reduced if the foreign tissue is first cleared of its resident presenting cells, even though the host's presenting cells should be able to pick up the foreign MHC molecules and activate T cells against them. Thus there is clearly a difference between a foreign MHC molecule in its original presenting cell membrane and the same molecule when it is picked up and displayed by the host's presenting cell. Perhaps the represented foreign MHC is no longer appropriately linked up inside the cell to act as a triggering device. It would then be treated like any other foreign molecule, activating only those T cells which recognize it in association with the host's own MHC proteins. Since the grafted tissue does not express these host MHC molecules, it would not be recognized by the activated T cells and would not be rejected.

What is it about their structure that allows MHC molecules to associate with so many foreign molecules? How is their structure related to their ability to act as triggering devices? Are T cells evolutionarily pre-programmed to look at them or are the specific T cells selected from a randomly generated population of resting cells? How many different MHC molecules are there? With the molecules and their genes in hand (see the following article) we may soon have the answers to these questions. □

Polly Matzinger and Rose Zamoyska are in the Department of Pathology, University of Cambridge.

0028-0836/82/250628 01$01.00

The evolutionary past of the major histocompatibility complex and the future of cellular immunology

from Miranda Robertson

THE penetration of the major histocompatibility complex (MHC) with cloned DNA probes has opened up a new approach to an old evolutionary problem — how to account for the unparalleled polymorphism of the cell-surface antigens it encodes This spring, representatives of the main laboratories involved met in Oxford to compare notes, and to discuss their strategies for research on the functions of the MHC products*

It is now generally accepted that in the remote evolutionary past, three of the four most prominent families of proteins of the immune system must have evolved from a single immunoglobulin-like domain The largest and most complex of the three related families — that of the immunoglobulin genes — is now scattered over three different chromosomes, the other two — the so-called class I and class II MHC antigens — are encoded together in the major histocompatibility complex (Fig 1), along with some of the fourth family — the components of complement, which so far seem to be structurally unrelated to any of the others The sequences of cloned MHC DNA have strongly reinforced the evidence for the common ancestry of the immunoglobulins and the class I and class II antigens, the present state of knowledge on the structure of the MHC genes, their products and their relationships with immunoglobulin genes are summarized in Fig 2 But most of the discussion at Oxford focussed on the much more recent evolutionary history of the MHC, and in particular on the polymorphism of the class I and class II antigens

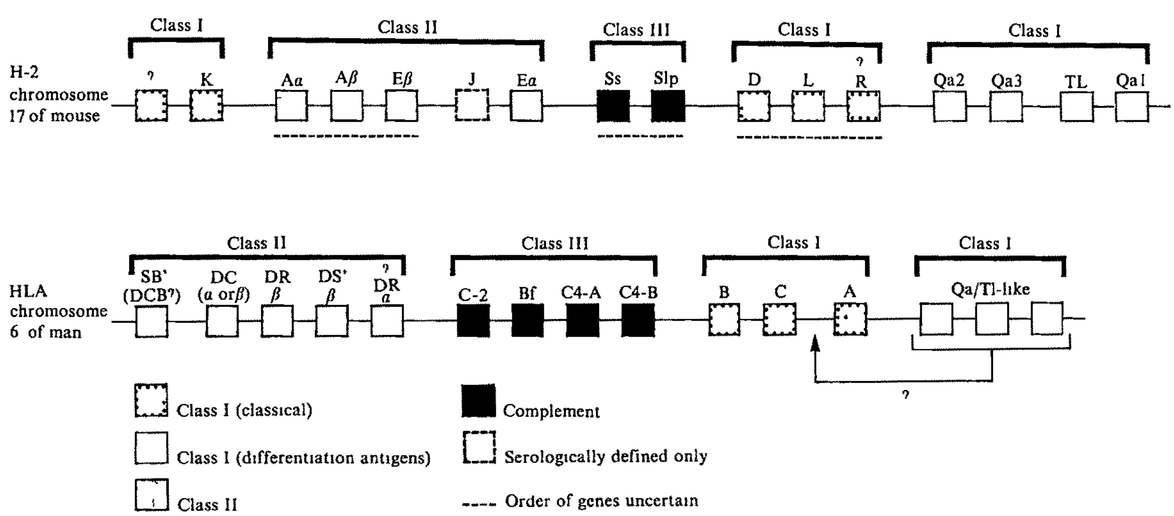
According to genetic and biochemical analysis, there are 30–60 alleles at each of the polymorphic loci encoding the class I antigens, and perhaps 12–40 at each of the class II loci Why and how has this extraordinary polymorphism been maintained in mammalian evolution? An answer began to emerge with the relatively recent discovery that the MHC antigens guide and regulate the immune response (see legend to Fig 1) It is known that the T lymphocytes that kill virus-infected cells, as well as those that regulate the responses of other lymphocytes to antigens, can recognize foreign antigens only in association with MHC molecules (see preceding article) It has also been possible to show that a given foreign antigen may be more efficiently recognized in association with some MHC alleles than with others Thus an animal heterozygous for MHC antigens may respond efficiently to a wider range of pathogens than a homozygote, and poymorphism may be maintained by heterozygous advantage The polymorphism may be stable, or the MHC molecules may have to evolve rapidly to keep pace with pathogens that are continually evolving ways to evade recognition, so that advantageous new variants are continually selected A deeper understanding of the selective pressures at work on MHC molecules will require a much better appreciation of the mechanisms of recognition and destruction of pathogens, and this is likely to be the reward of one of the new techniques discussed at Oxford For it is now possible to transform cultured cells with cloned MHC genes, and the use of manipulated genes in such experiments may settle these

*A meeting on the 'Cloning of the HLA and H-2 Regions', organized by W F Bodmer and sponsored by the Imperial Cancer Research Fund and the European Molecular Biology Organization, was held in Oxford on March 21-24 1982

Miranda Robertson is associate editor of Nature

Fig 1 The major histocompatibility complex (MHC) of mouse and man The MHC molecules can be divided into three classes on the basis of their structure and function The class I antigens constitute a single class structurally (see Fig 2*a*), but fall into two functional groups The first of these contains the "classical" class I antigens, first discovered as the transplantation antigens and now known to function as target antigens in the recognition and destruction of virus-infected cells by cytotoxic T lymphocytes They are expressed on virtually all somatic cells The second group of class I antigens, loosely defined as lymphocyte differentiation antigens because of their tissue distribution, have no known function The class II antigens are expressed largely on B lymphocytes and macrophages of the immune system, and are believed to be essential for presenting antigen to the helper and suppressor T cells that regulate the immune response The class III products are components of the complement system These maps are based on serological and biochemical data it is now known that the MHC contains other class I sequences not detectable serologically Moreover, not all mice express all the genes indicated on the *H-2* map some strains are known to lack a functional *L* gene, and the evidence for a second *K* gene is extremely indirect[1] The *J* gene is the most problematic of all because its product has eluded all attempts at biochemical identification Because its determinants have been detected on candidate T-cell receptor molecules and antigen-specific factors, the *J* product has become something of an immunological holy grail Now that Steinmetz has a probe for the neighbouring Eα gene (see text), he proposes to walk down the chromosome to the *J* locus this should settle the question of whether there is anything there (disputed by Klein *et al*[2]) The exact distance he will have to walk is not certain

In the mouse, different groups of alleles have been segregated into inbred strains, each linked group (haplotype) being designated by a superscript letter Thus B 10 strain mice, for example, have the *b* haplotype, and so bear antigens H-2K^b and so on



0028-0836/82/250629 04$01 00

and other fundamental questions that have resisted the efforts of cellular immunologists for many years

But in the meantime, sequence analysis of the cloned DNA more directly addresses the question of how — rather than why — the polymorphism has arisen Immunogeneticists have speculated, for example, that the alleles may not be alleles of a single gene at all, but tightly linked clusters of different genes[7,8] This is now ruled out The class I genes do occur in clusters of related sequences, at least in mouse[9], but although there are at least twice as many class I sequences as would be predicted from genetic analysis, there are not enough to account even for a substantial fraction of the serologically detectable alleles The extra class I sequences thus pose a fresh evolutionary problem — particularly in the light of preliminary evidence that the class II antigens, by contrast, have few or no supernumerary relatives in the genome The nature and significance of the extended class I gene family was one of the main topics of discussion at Oxford

Evolution of the class I genes

The class I genes seem to constitute a large and complex family is a state of evolutionary flux In all genomes for which DNA probes are available, there are at least twice as many sequences hybridizing with class I probes as there are class I genes on the genetic map Pigs (D Singer, US National Institutes of Health, Bethesda) and men[10] probably have 15–20, and mice have as many or more, depending on strain The largest and most complex class I family discovered so far is that of the BALB/c mouse, which has 36 hybridizing sequences organized in 13 clusters[9], the B 10 mouse, by contrast, may have only 15–20 (R Flavell, National Institute for Medical Research, London), and comparisons between restriction maps from other strains reveal yet more variation[9]

This evidence for a history of recent duplications and deletions is strongly reinforced by two features of the sequence clusters first, more homologous genes seem to be clustered together, and second, the noncoding DNA flanking the clustered genes is often strikingly homologous[11] According mainly to restriction mapping, the duplications seem to be concentrated largely at the *Qa/TL* end of the mouse MHC (M Steinmetz, Caltech, Flavell, D Schulze, Albert Einstein Medical School, New York), and preliminary analyses of human cell lines[10] suggest a similarly skewed distribution in man

Now that the sequencing of genomic class I DNA has begun, it is becoming clear that a substantial proportion of the class I family may be pseudogenes Of the three BALB/c genes that have been sequenced, two encode functional L^d antigens[12-14] and one is a pseudo-*Qa* gene Of four human class I genes sequenced (or partially sequenced), one encodes an A2 antigen, one a B7, one is clearly a B pseudogene and the fourth may be another pseudogene (A Biro, Yale University, and ref 15)

Why should so many cousins of the class I genes — some evidently nonfunctional — have been maintained in the mammalian genome if they do not contribute to MHC antigen polymorphism? The answer favoured by most of those working on them is that in fact they do, but indirectly W Bodmer (Imperial Cancer Research Fund, London) has speculated that variant proteins might be produced by differential splicing of primary transcripts containing more than one gene There is so far no evidence for this On an evolutionary timescale, L Hood (Caltech) has pointed out that polymorphism could be increased by the tendency of partially homologous sequences to recombine, in particular by unequal crossing-over[9] Others (P Kourilsky, Pasteur Institute, and refs 12 and 16) propose, more specifically, that the genes and pseudogenes at the *Qa* and *TL* regions of the mouse may represent 'reservoirs of polymorphic variation' that can be tapped by means of gene conversion Since the nature and implications of the evidence for gene conversion attracted more discussion than any other speculation on class I gene evolution, it is worth recapitulating briefly how it is believed to work

Gene conversion, originally found in fungi, is a mechanism by which genetic information can be transferred from one gene to another related gene anywhere in the genome — although allelic genes and members of tandem families are the most likely recipients It is believed to occur through pairing between partly homologous sequences during meiosis or mitosis, followed by mismatch repair resulting in the conversion of one sequence to the other There have recently been increasing indications of such cross-talk between different members of gene families in the course of vertebrate evolution, the evidence for gene conversion in the MHC rests chiefly on the discovery of clusters of substitutions, detected both at the protein and at the DNA level, in individual genes

For example, J L Strominger (Harvard University) finds a cluster of nine residues in which the *HLA-A2* allele differs from the *HLA-A28* allele but is identical to the non-allelic *HLA-B7* Gene conversion is one way of explaining how clustered differences between alleles could be identities between related tandem genes Still at the protein level, S Nathenson (Albert Einstein Medical School, New York) quoted the case of 'mutant' mice from inbred strains showing multiple substitutions in MHC antigens[17] — a pattern that is not consistent with point mutation and has the additional and highly suggestive peculiarity that the same substitutions seem to occur repeatedly in different mice

Some preliminary comparisons of mutant protein sequences and recently cloned genes now strongly suggest that the mutant sequences derive from other genes D Margulies (US National Institutes of Health, Bethesda), for example, comparing a cloned *H-2L^d* gene from a BALB/c mouse with the H-2K^b antigen of the B 10 mouse, and with mutant H-2K antigens, finds that in 9 out of 10 cases, differences between wild-type and mutant K^b sequences are identities between the mutant and H-2L sequences Further eccentric relationships between *H-2* loci are revealed by a comparison of the same *H-2L* gene with other class I sequences they show that the BALB/c *H-2L^d* gene is more closely related to the *H-2D^b* gene of the C57BL/6 mouse than to other class I genes of its own strain[12] There are two possible explanations for this The first is that an exchange of information took place between an *H-2L^d* gene and an *H-2D^b* gene before the development of the inbred strains that segregated all the *H-2^d* genes into BALB/c and all the *H-2^b* genes into C57BL/6 Alternatively, the *H-2L^d* gene may have been converted by an unexpressed pseudogene within the BALB/c MHC

In the light of these possibilities, both Margulies and Flavell are focusing their attention particularly on an *H-2K* mutant, *H-2K^bml*, that arose within the B 6 mouse strain H-2^{bml} contains two amino acid substitutions that entail at least four base changes (so that point mutation is unlikely), and are homologous to the BALB/c *H-2L^d* gene Because the mutation occurred within the B 6 strain, it cannot be due to conversion by an *H-2^d* chromosome, and may be due to conversion by an *L^b* gene or an *L*-like pseudogene in the B 6 chromosome This hypothesis can be checked by searching the B 6 MHC with the appropriate probe

Gene conversion is in fact also an alternative explanation for the strong homologies seen in the flanking regions and introns of MHC genes, since conversion would be expected indiscriminately to homogenize coding and moncoding DNA Indeed, the difficulty of distinguishing between crossing-over and conversion as the mechanism underlying any of the genetic exchanges in the MHC was repeatedly stressed by W Bodmer (Imperial Cancer Research Fund, London) To produce a cluster of subsitutions with no change in gene number it would be necessary to postulate double unequal crossing-over, which may seem cumbersome and improbable, but there is no way in which to distinguish formally between that and gene conversion in higher animals (For a particularly lucid exposition of the effects of crossing-over and conversion on gene copy number, see ref 18) The discovery of gene conversion in fungi was possible only because in some species the products of a single meiosis remain in a tightly clustered tetrad in which any departure from mendelian ratios is

directly detectable It was the occasional observation of tetrads in which the expected 2 2 ratio for a given allele proved to be 3 1 that revealed the occurrence of gene conversion between alleles

Such precise analysis is not possible in mice, in which each off spring in a litter is probably the product of a quite different meiosis, and the mendelian ratio for a given litter cannot be expected to be exact The stability of mendelian ratios averaged over large numbers is, however, notorious, which shows that gene conversion must be rare, and raises a further problem, also emphasized by Bodmer While various mechanisms can account for the creation of pseudogenes and the conversion of functional ones, it may be much more difficult to account for the spread and fixation of these rare genes in the population Selection does not anticipate, so the pontential advantage of pseudogenes as a bank of variants could not drive a non-functional gene to fixation There are various general ways of explaining the fixation of genes in the, absence of selective pressure one is through the founder effect, in which a large population grows from a few individuals derived from a rare mutant, another is the hitch-hiker effect, in which the event that creates a nonfunctional gene or a mutation also produces an effect in a closely linked locus that confers an immediate selective advantage The non-mendelian mode of inheritance of converted genes may be very important in this connection, since it could accelerate spread provided that conversion were more probable in one direction than in the other

But a more specific, and much more controversial possibility was raised by G Jay (US National Cancer Institute, Bethesda), who presented data implying that at least some pseudogenes may not be pseudo at all, but coding sequences for secreted products

When is a gene not a gene?

The basis for Jay's suggestion is a cDNA clone derived from the SWR/J mouse[19] and whose sequence is broadly homologous to those of the *H-2K* and *H-2D* genes up to the transmembrane region Here, through a series of substitutions and deletions, it encodes hydrophilic residues instead of hydrophobic ones, and moreover contains a stop codon The cDNA was originally synthesized by reverse transcription from liver cell mRNA, and Jay now reports that liver is the only tissue in which the gene is transcribed, thus the tissue distribution of its product would be quite unlike that of the known *H-2K* and *H-2D* products, which are ubiquitous This, and some recent (and rather controversial) evidence that liver transplants may suppress graft rejection[20], led him to propose a function for the product in the induction of immune tolerance At this stage, however, there is only indirect evidence that the transcript is actually translated or secreted Jay finds a protein of the right size in serum, but it has not been fully characterized and is not known to be secreted by liver cells

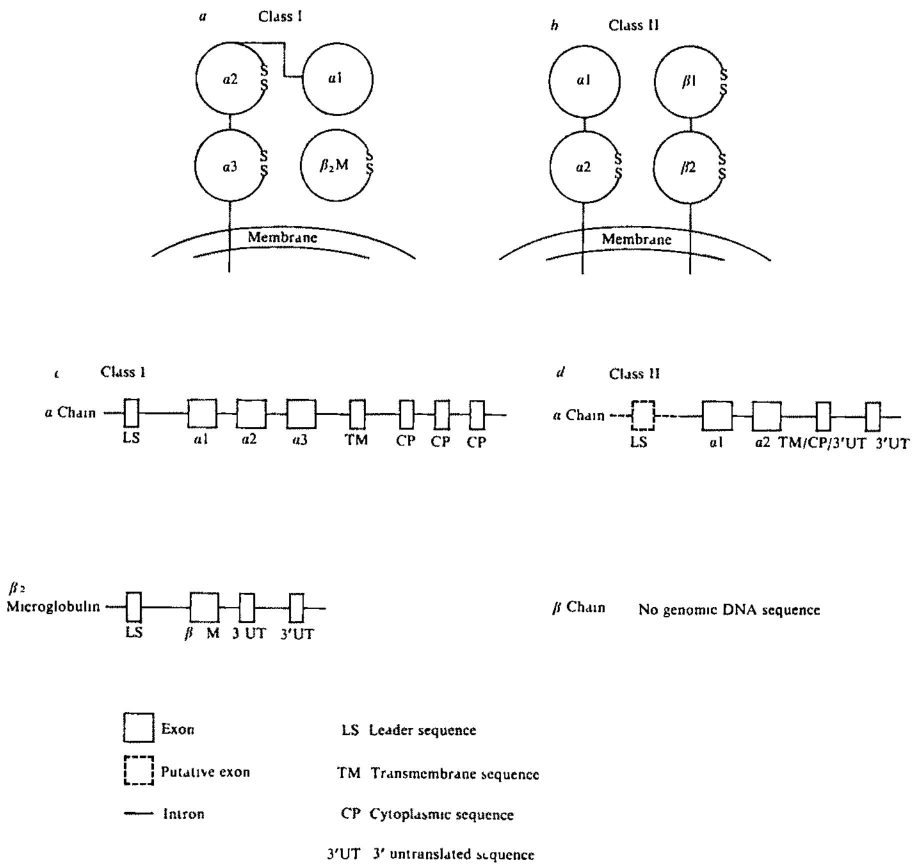


Fig 2 Schematic diagrams of the class I (*a*) and class II (*b*) antigens, drawn so as to emphasize the probable homology of their quaternary structures Each molecule has four domains $\alpha1$, $\alpha2$, $\alpha3$ and β_2 microglobulin for class I molecules, and $\alpha1$, $\alpha2$, $\beta1$ and $\beta2$ for class II molecules *a*, Class I antigens are composed of an α-chain that in the case of the classical antigens is very highly polymorphic The α-chain spans the plasma membrane, but can be expressed on the cell surface only in association with β_2-microglobulin, a highly conserved molecule encoded outside the MHC and indeed on a different chromosome (2 in mouse, 15 in man) β_2-microglobulin, which does not span the membrane, was early recognized as having considerable homology in general domain structure with immunoglobulin (it has been described as an "orphaned immunoglobulin domain") More recently, the $\alpha3$ domain of the β-chain has also been shown to be homologous to immunoglobulin[3] *b*, The two chains of the class II antigens are both encoded in the MHC, and both span the plasma membrane The α-chain is relatively non-polymorphic, the β-chain highly polymorphic Homology to immunoglobulin has been demonstrated in the $\alpha2$ domain (A Korman, Harvard University D Larhammar, Uppsala University) and in the $\beta2$ domain[4,6] The domain boundaries of the proteins correspond to the exon-intron boundaries of the DNA of all chains for which genomic clones are available (*c* and *d*)

Although for the time being participants at the Oxford meeting were inclined to be sceptical about the proposed function of the liver transcript, it did lead to a reexamination of the criteria used to identify pseudogenes, and it focused attention on some interesting features of the 3′ end of the MHC gene sequences

A pseudogene can be identified by (1) the deletion of coding sequences, (2) the presence of termination codons in coding sequences, and (3) non-conservation of the RNA splicing signals at intron–exon junctions The pseudo status of the *HLA* gene sequenced by Biro, which in any case has a frameshift mutation in the first exon, is by these criteria unassailable The mouse sequence of Steinmetz *et al*[11] is a more ambiguous case Its departures from orthodoxy do not begin until its transmembrane exon, which encodes one charged residue that ought to be hydrophobic, and contains a termination codon The remaining mutations — another stop codon and a non-conserved splice site — occur 3′ to the transmembrane exon, so it is not impossible that this pseudogene encodes a secreted product

A question of paramount importance is thus how cloned genes can be assayed quickly for function This question was raised by Hood in the context of an interesting drawback that has just come to light in the use of transfected cells for this purpose The technique depends on the ability of transfected mouse L cells to express the cloned gene on their surfaces, where the product can be detected serologically[12,13] or indeed function as a target for *H-2*-restricted virus-specific cytotoxicity[21] However, Goodenow now

reports that transfection of L cells with cloned DNA containing only the 5′ half of a class I gene has resulted in the surface expression of a new class I molecule How the incomplete gene contrived to pick up the necessary 3′ half containing the transmembrane sequences is a mystery To avoid being misled by such rescued genes in the future, Hood suggests sequencing the most variable regions of cloned genes, synthesizing the corresponding peptides, raising monoclonal antibodies against them, and using the antibodies to test for expression on tissue cells While this would not abolish false positives (because of the possibility of cross-reacting determinants), it should reduce them (Not everyone, of course, has access to all the necessary facilities)

But the most exciting application of transfected L cells will be in the elucidation of MHC antigen functions by reverse genetics Both Hood and Flavell gave notice of their intention to manipulate the sequences and reshuffle the exons of cloned genes in order to establish what these manoeuvres will do to the functions of their products

Of particular interest is the part played by the cytoplasmic domains in immune function, because (for example) if the recognition of an MHC antigen triggers a signal from the recognized cell, it must be the cytoplasmic domains that mediate the release of the signal Furthermore, there are hints of alternative splicing or termination sites at the 3′ end of the MHC sequences Malissen *et al*[15] have suggested, rather indirectly from comparisons of human genomic with mouse cDNA, that the pattern of splicing of the three cytoplasmic exons may vary — an idea that is now supported by comparisons with human cDNA (J Trowsdale, ICRF), and Steinmetz *et al*[11] have evidence that the last intron may not always be removed from mouse transcripts of class I genes There is no easy way to check what is normally transcribed from most of the genomic clones, because the probes used to obtain them are in many cases from a different gene or even species This also makes it difficult to check for cloning artefacts that might account for some of the unexpected features of the genes To resolve these questions, it will now be necessary to go back and synthesize cDNA from the messengers transcribed from the cloned genes themselves in the cells that normally express them

Class II genes and conservation

Cloning of the class II genes has only just begun, but it is clear even at this stage that unlike the class I genes they correspond more nearly in number to their known products Probes for the α-chain genes have consistently revealed only one strongly hybridizing band in man (A Korman, Harvard University, H Ehrlich, Cetus, D Larhammar, Uppsala University, J Lee, Imperial Cancer Research Fund, London, C Wake, Geneva University) and in mouse (Lundt, NIMR, London, Steinmetz, Seidman) The strongly hybridizing band in man has been detected using a *DRα* gene, which is homologous to mouse *IEα* Three additional bands have been detected by Trowsdale under relaxed hybridization conditions One of these may be the *DC1* gene — the putative human homologue of the mouse *IAα* it is present in a different genomic clone from that containing the *DRα* gene, and may be homologous to a cDNA clone identified by preliminary sequence analysis as *DC1* (C Auffray, Harvard University) The identity of the third band is unknown There is some uncertainty about how many human class II α genes to expect, so three would not be a major surprise But two surprises have been delivered by human α-chain clones Korman and Lee have shown that there are two polyadenylation signals in the 3′ untranslated cDNA sequence, which may explain preliminary evidence from Lee for *DRα* mRNAs of different lengths The significance of this finding is not known

The second surprise is a *DRα* cDNA clone containing a termination codon before the hydrophobic sequence (Larhammar) This could be interpreted as corresponding to a transcribed pseudogene, or to a real gene encoding a secreted product, or it may be a cloning artefact The last possibility has to be taken particularly seriously in this case, because of the small size of the *DRα* family

The absence of large numbers of extra α loci may not be entirely surprising, in the light of the preceding evolutionary speculations, since the α chains of the class II antigens are relatively non-polymorphic But the β-chains may have as many as 20–40 alleles per locus, yet so far neither man (L Rask, Uppsala University, E Long, Geneval University, and ref 22) nor mouse (Lundt, National Institute for Medical Research, London, L Clayton, Massachusetts Institute of Technology) seems to have more than two or three hybridizing sequences If polymorphism depends on frequent crossing-over, or a reservoir of variant sequences for gene conversion, why is the class II gene family not as large as the class I family? There are two possible answers The preliminary estimates on gene number may be wrong (evidence from protein sequences suggests six to eight genes, rather than two or three in man[23]), or hybrids between different α- and β-chains may conceivably make a significant contribution to phenotypic polymorphism, obviating the need for so large a genetic source

It is, however, worth noting that not everyone would expect the high polymorphism of the MHC to be associated with rapid evolution J Klein has argued from genetic analyses of wild mice that MHC polymorphisms are extremely stable in evolution, that neither mutation nor recombination is frequent, and indeed that the different allelic variants of the MHC antigens may have diverged before mouse diverged from man[24] In short, the mammalian MHC must have undergone a period of expansion and divergence early in its evolution, but long ago have become maximally adapted to deal with the prevailing pathogens It may therefore be necessary to look elsewhere for an evolutionary explanation for the structure of the MHC

One possibility that has to be faced is that it may have no adaptive significance at all, and merely represents a tableau of the processes everyone has assumed must underlie protein evolution It is generally accepted that families of structurally related but functionally distinct genes must have evolved by duplication and divergence (as indeed so many of the genes of the immune system presumably evolved from a privaeval immunoglobulin-like domain), and there is every reason to suppose that during the period of divergence one of the duplicated genes would have become non-functional, through an accumulation of the termination codons, frameshifts and deteriorated splicing signals that characterize pseudogenes Gene conversion offers a mechanisms for correcting many such disabling mutations in one step — thus enabling the diverging gene to be 'tested' at intervals for new functions[25] The random introduction of new stop codons or poly(A) addition sites could produce shorter or longer molecules, or transform a membrane-bound to a secreted molecule if they occurred before a region encoding hydrophobic residues Since we assume that such processes are random, we should not necessarily expect every transcribed or even translated gene to contribute to survival

1 Demant P and Ivanyi D *Nature* **290**, 146 (1981)
2 Klein, J, Juretic, A Baxevanis C N & Nagy Z A *Nature* **291**, 455 (1981)
3 Orr, H T *et al* *Nature* **282**, 266 (1979)
4 Larhammar D *et al* *Scand J Immunol* **14** 617 (1981)
5 Kaufman, J & Strominger J L *Nature* in the press
6 Larhammar D *et al* *Proc natn Acad Sci U S A* in the press
7 Bodmer W F & Bodmer J G *Brit med Bull* **34** 309 (1978)
8 Silver, J & Hood L E *Proc natn Acad Sci U S A* **73**, 599 (1976)
9 Steinmetz M Winoto A Minard K & Hood L *Cell* **28**, 489 (1982)
10 Orr H T *et al* *Nature* **296** 454 (1982)
11 Steinmetz M *et al* *Cell* **25** 6783 (1981)
12 Evans G A *et al* *Proc natn Acad Sci U S A* **89**, 1994 (1982)
13 Goodenow R, *et al* *Science*, **215** 677 (1982)
14 Moore K W *et al* *Science* **215** 679 (1982)
15 Malissen, M, Malissen, B & Jordan, B R *Proc natn acad Sci U S A* **79**, 893 (1982)
16 Flavell, R A *et al* ICN/UCLA Symposium on Molecular & Cell Biol **26** (edit by O Malley B & Fox F C) (Academic in the press)
17 Nairn R Yamaga K & Nathenson, S G *Ann Rev Genet* **14**, 271 (1981)
18 Baltimore D *Cell* **24**, 592 (1981)
19 Cosman D Khoury G & Jay G *Nature* **295**, 73 (1982)
20 Kamada N, Davies, H & Roser B *Nature* **292** 840 (1981)
21 Hood L *et al* *Nature* **297** 415 (1982)
22 Wiman, K *et al* *Proc natn Acad Sci U S A* **79**, 1703 (1982)
23 Kratzin H *et al* *Hoppe Seylers Zeit Physiol* **10** 362 1665 (1981)
24 Arden, B & Klein J *Proc nant Acad Sci U S A* **79**, 2342 (1982)
25 Jeffreys A in *Genome Evolution* (edit by Dover G A & Flavell, R B), 157 (Academic London 1982)

REVIEW ARTICLE

Repetitive sequence transcripts in development

Eric H. Davidson & James W. Posakony

Division of Biology, California Institute of Technology, Pasadena, California 91125, USA

Interspersed repetitive sequences are represented widely in animal cell nuclear RNAs, in the poly(A) RNA stored in eggs and in some mRNAs. Their expression is developmentally modulated. Although the genomic location of repetitive sequences may change rapidly during evolution, the patterns of their transcription suggest a variety of possible functions.

THE repetitive DNA sequences of the animal genome are extensively represented in cellular RNA. The patterns of repeat sequence expression have been known for some years to vary during development, and from tissue to tissue[1]. However, until cloned probes became available, it was usually not possible to investigate specific repeat sequence families, or to distinguish quantitative developmental changes in repetitive sequence expression (that is, changes in the number of repeat transcripts of given sequence families) from qualitative changes (differences in the identity of the repeat sequence families represented in RNA). The availability of repetitive sequence clones, each representing a single family of more or less homologous sequences, has now opened new areas of investigation. The typical animal cell contains several very different kinds of repetitive sequence transcript, including (1) mRNAs derived from families of related structural genes, such as the histone or actin mRNAs; (2) the small nuclear RNAs (snRNAs), usually only a few hundred nucleotides or less in length and transcribed from members of certain specific repetitive sequence families; and (3) heterogeneous nuclear RNAs (nRNAs) and some cytoplasmic poly(A) RNAs that include repeat sequence transcripts. The latter RNAs are generally relatively large [1 to >10 kilobases (kb)], and are believed to be transcribed by polymerase II. This brief review is focused mainly on this heterogeneous class of repeat transcripts, the very existence of which remains a major puzzle of molecular biology.

Interspersed organization and diversity of transcribed repetitive sequences

The population of RNAs present in the cell nucleus has a predominantly interspersed sequence structure. Covalent linkage of various short repeat sequence elements, typically only a few hundred nucleotides long with transcripts of single copy sequences from the same genomic transcription unit has been shown for the nuclear RNA of HeLa cells[2–5], sea urchin embryos[6] and rat ascites cells[7]. An interspersed sequence organization is also characteristic of the maternal poly(A) RNA stored in sea urchin[8,9] and amphibian[10] eggs. The sea urchin egg poly(A) RNA transcripts that include repetitive sequences average 5–10 kb in length and consist of repeat elements a few hundred nucleotides long, interspersed with sequences hybridizing to single copy DNA (ref. 9 and J.W.P. *et al.*, unpublished data). Both complements of many or all of the transcribed repetitive sequences are represented in the interspersed RNAs, though generally in different molecules. Therefore the egg poly(A) RNA molecules can be renatured to form branched multimolecular structures that can be visualized in the electron microscope (Fig. 1*a*).

Although the repeat-containing poly(A) RNAs of sea urchin and *Xenopus* eggs include single copy or near-single copy sequences later found on embryo polysomes[9,10], it is not yet clear whether any of the repeat elements themselves are included in polysomal embryo mRNAs (see, for example, refs 9, 11–15). Interspersed repeat transcripts have been demonstrated unequivocally in the mRNA of the cellular slime mould *Dictyostelium*[16–18]. Firtel and his associates identified two specific families of short repeat sequences at the 5′ ends of sets of mRNAs, most of which consist of otherwise unrelated single copy transcripts[16,17]. However, in at least one of these *Dictyostelium* mRNA families, repeat sequences are represented only in an asymmetric fashion, so that although the transcribed repeats are dispersed in the genome only one complement is found in the mRNA[17]. There is also evidence that in human lymphoblastoid cells, about half the polysomal poly(A) RNA contains repeat sequences, some of which may belong to the high-frequency *Alu* repeat family[19]. The *Alu* family[20] is also expressed prominently in HeLa cell heterogeneous nRNA[21]. Members of the *Alu* repeat sequence family apparently contain promoters for polymerase III and are represented in a major family of snRNAs[22–25], the 7S snRNAs.

Almost nothing is known of the diversity of the transcribed repeats represented in the interspersed RNAs of human and other mammalian cells. Most recent studies have focused on only one of the large number of repetitive sequence families present in these genomes (J. W. Roberts, R. Hudspeth, E.H.D. and R. J. Britten, unpublished data and refs 26, 27), the *Alu* repeat family or (in other species) its close relatives. However, several different repeat sequence families are represented in the poly(A) RNA of human lymphocytes[28]. Direct measurements[8,9,29] show that in the sea urchin, at least a few hundred different repeat sequence families are represented prominently in egg poly(A) RNA. Of importance in considering the significance of repeat sequences is the recent finding that only a few members of each family may actually be transcribed to any extent (J.W.P. *et al.*, unpublished data). Figure 1*b* displays sea urchin egg and embryo RNA blots carried out with the separated strands of a cloned probe containing a typical short repetitive sequence. Each of the several transcript species visualized derives from a different member of the large repeat sequence family represented by the probe. Both strands of given repeat sequences occur in the interspersed RNAs, as indicated in parts *a* and *b* of Fig. 1, because some of the diverse transcription units including these repeats are oriented in one direction with respect to these particular repeat sequences, while others are oriented oppositely.

Developmental regulation

Complex as it is, the pattern of interspersed repeat sequence transcription is by no means constitutive and unchanging. Comparison of the second and third tracks in Fig. 1*b* reveals a qualitative developmental change in the expression of the

0028-0836/82/250633—03$01.00

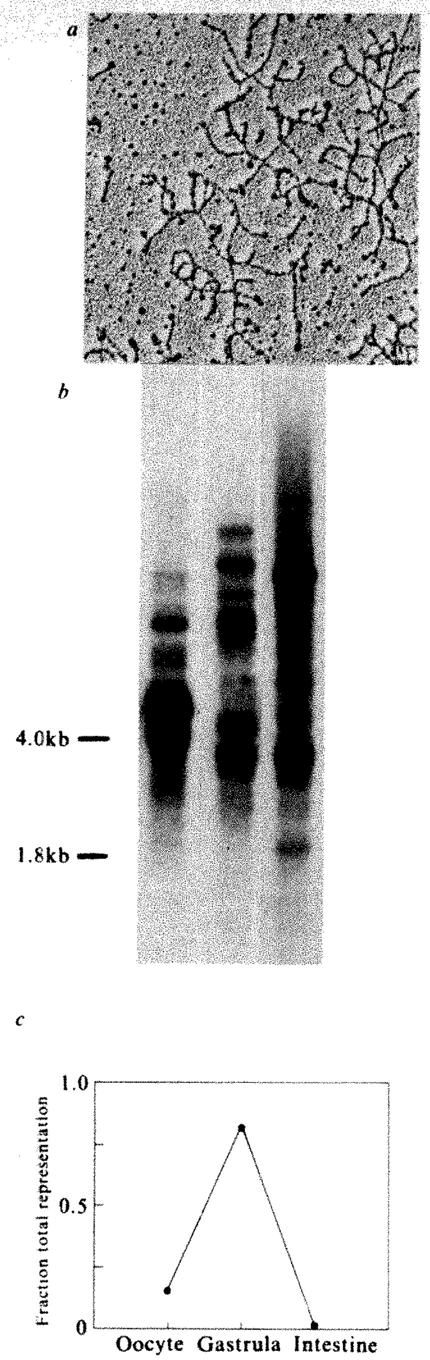


Fig. 1 Characteristics of egg and embryo poly(A) RNAs containing interspersed repetitive sequences. *a*, Poly(A) RNA of mature *Xenopus* oocytes, renatured to RNA C_0t 600 and spread for electron microscopy from 80% formamide[10]. A distance of 1 cm represents a single-stranded RNA length of 870 nucleotides. The branched mutimolecular RNA networks are held together by base-paired repetitive sequence duplexes, and include about 70% of the total oocyte poly(A) RNA. Fedoroff *et al.*[4] showed earlier that structures similar to these are formed when nuclear RNA is renatured. *b*, RNA gel blots using separated strands of a short interspersed repetitive sequence clone called CS2109B. Left, 'L' strand reacted with egg poly(A) RNA; centre, the complementary 'U' strand reacted with the same RNA; right, the same strand as used in the centre track reacted with blastula stage embryo poly(A) RNA; the exposure of the blastula track is several-fold lower (unpublished experiments of J.W.P., R. J. Britten and E.H.D.). *c*, Relative extent of expression ('per cent representation'[30]) of the 'U' strand of the 2109B repeat sequence in gastrula nuclear RNA and adult intestine cell nuclear RNA, compared with egg RNA, from molecular titrations carried out with the cloned probe[30].

repeat family, which implies the transcription in blastula stage embryos of additional members of the family. Scheller *et al.*[30] found for each of nine sea urchin repeat clones a unique pattern of representation in embryo nuclear RNA, which differed in adult intestine cell nuclear RNA and again in egg RNA. Representative data are shown in Fig. 1*c* for the same clone as used for the RNA gel blots in Fig. 1*b*. In fact, the embryo and adult intestine nuclear RNAs of the sea urchin are much more easily distinguished by their differing contents of specific repeat sequence transcripts[30] than by their largely overlapping sets of single copy sequence[31–33].

A second system in which developmental regulation of repeat sequence transcripts has been observed is *Dictyostelium*. One of the mRNA sets sharing a common repeat sequence[17] is very rare in vegetative cells but becomes increasingly prevalent as development proceeds, beginning at ~5 h (A. R. Kimmel and R. A. Firtel, personal communication). RNA gel blots carried out with single copy probes obtained from two of these transcripts show that both appear at about the same time. The developmental change in the abundance of this set of mRNAs is about 10-fold. Zuker and Lodish[18] also reported developmentally regulated sets of interspersed cytoplasmic *Dictyostelium* poly(A) RNAs, each of which is defined by the presence of a particular repeat sequence. At least some members of these sets are 'coordinately regulated', to the extent that in one case, the multiple transcripts cannot be detected in vegetative cells but appear by 5.5 h of development, while another set appears only by 15 h.

Developmental regulation of a different kind of repeat sequence transcript has been reported in *Drosophila*. These transcripts are exemplified by the poly(A) RNAs carrying the *copia* sequence. The *copia* repeat sequences are about 5 kb long, and are apparently able to transpose at a relatively rapid rate, as their location in the genome as well as their number differ among *Drosophila* strains and among tissue culture cell lines[34,35]. Structural features of the *copia* element and its sites of insertion are reminiscent of other eukaryotic and prokaryotic transposable elements[36–38]. The *copia* sequence is transcribed into cytoplasmic poly(A) RNAs ~2 and 5 kb long, and at least the 2 kb RNA can be translated, and thus, is thought to be an mRNA[38]. The *copia* transcripts differ basically from the repeat-containing RNAs thus far mentioned in that they do not have an interspersed sequence structure. The 2 and 5 kb transcripts begin and end within the *copia* repeat sequence element[35]. Flavell *et al.*[38] and Scherer *et al.*[39] have reported that the level of *copia* transcripts is modulated during development. These transcripts are not detectable in early embryos but appear after 10 h, and they are most concentrated in larvae. Developmental patterns of regulation have also been reported for transcripts of at least three other *Drosophila* long repeat sequence families that appear to be transposable and to share the general features of the *copia* sequences[39,40].

Are repeat sequence transcripts biologically significant?

The observation that repeat sequence expression is correlated with developmental change does not, of course, demonstrate function, though repeat sequences might seem much less interesting were their transcription non-existent or invariant. Clearly there are diverse kinds of developmentally regulated repeat sequence transcript, the properties and potential significance of which must differ in basic ways. The 5′ location of the repeats and their asymmetric representation in the *Dictyostelium* mRNA sets studied by Kimmel and Firtel[17], and the apparent coordinate expression observed both for these transcripts and the mRNA sets studied by Zuker and Lodish[18], led both groups to suggest that *Dictyostelium* mRNA repeat elements serve as coordinate regulatory sequences[41]. Such arguments are unlikely to pertain to *copia*-like repeats because these elements mainly promote transcription of their own sequence. On an evolutionary scale, of course, the presence of transposable elements that

include promoters could be of fundamental importance, because readthrough transcription of adjacent genes might occur as the result of an insertion event. A very instructive example is found in the yeast ROAM mutants studied by Errede *et al.*[42], who concluded that insertion of a Ty-1 transposable repeat element (and perhaps additional flanking sequences) in the 5′ region of the iso-2-cytochrome *c* gene causes this gene to be brought under the positive control of diffusible regulators from the mating-type locus.

There are two opposing attitudes towards the significance of the interspersed short repeats lying in and around the transcription units of animal genomes[43–45]. As proposed originally by Britten and Davidson[45], interspersed repeats seem to 'diffuse' around the genome during evolution. Recent relevant evidence includes the findings that the family sizes of specific interspersed repetitive sequences differ strikingly between related sea urchin species[46]; the mammalian *Alu* repeat sequences and their homologues are flanked by brief terminal repeats similar to those that surround known transposons[23]; and in human[47] and monkey[48] cells, these same sequences can be recovered from closed circular extrachromosomal DNA, just as *copia* sequences can be found in closed circular DNA of *Drosophila* cells[49]. Thus one way of interpreting the developmental modulations observed in interspersed repeat transcripts is simply to conclude that among the hundreds of thousands, or, in some genomes, millions of short repeat elements are some that happen to have been transposed, harmlessly, into developmentally modulated transcription units. Though functionless, these repeat elements would be transcribed merely because they lie in the path of the polymerase. Thus, for example, interspersed repeats are known to be present in intervening sequences of several genes, including those for vitellogenin[50], growth hormone[51] and conalbumin[52]. Such repeats would be included in the primary nRNA transcripts from these genes, and their expression would clearly appear developmentally modulated, because those genes are active only in certain cell types. The alternative view is that it may be premature to conclude that because molecular biology has not yet demonstrated a functional role for interspersed repeat transcripts, none exists in the cell. Hidden within the vast networks of sequence homology formed by dispersed genomic repeat families there might be small subsets of homologous sequence elements that serve as receptors for diffusible regulators and are involved in coordination of gene activity as proposed earlier[41]. The proposition that members of interspersed repeat families are undergoing a continuous, evolutionarily rapid redistribution suggests that many insertions and deletions are without phenotypic effect, and thus it is reasonable to suppose that even if some repeat elements of large families were to perform regulatory functions of some kind, most of their homologues are without function.

The interspersed transcripts stored in mature eggs are especially interesting from the standpoint of possible function, because they represent the major part of the mass of the egg poly(A) RNA, and result from accumulation during a long period of synthesis in oogenesis. They are evidently destined for use during embryonic development. These transcripts in some ways resemble nRNAs, although they are stored in the oocyte cytoplasm[10]. The repeat elements themselves are not likely to be translatable (ref. 29 and J.W.P. *et al.*, unpublished data) and could be located in long 3′ mRNA 'tails', or in internal positions such as intervening sequences that have not been processed out. Perhaps these interspersed maternal transcripts can be used in embryogenesis only after further modification, in contrast to the maternal mRNAs capable of immediate translation that are clearly also present in the egg[1]. Embryonic processing of maternal transcripts could be an important developmental regulatory event, or these transcripts could play other developmental roles altogether, as suggested by Costantini *et al.*[9]. Other recent speculations have centred on the intermolecular complementarity of the repeat elements in nRNA, and of some snRNAs with nRNAs. Intermolecular complementarity could provide the molecular basis for splicing or other processing decisions (for example refs 53, 54).

In summary, a variety of interpretations have so far resisted exclusion. Despite their ubiquity, their quantitative prominence, their apparent developmental regulation and the amount of interest they have aroused, the repetitive sequence transcripts of animal cells remain a phenomenon in search of a physiological meaning.

This work was supported by NIH grant HD-05753.

Note added in proof: Calvet *et al.*[55] have demonstrated by *in situ* psoralen cross-linking that at least U1 and U2 snRNAs are base paired to nRNAs within the cell nucleus.

1. Davidson, E. H. *Gene Activity in Early Development*, 211–218; 223–230 (Academic, New York, 1976).
2. Darnell, J. E. & Balint, R. *J. cell. Physiol.* **76**, 349–356 (1970).
3. Jelinek, W., Evans, R., Wilson, M., Salditt-Georgieff, M. & Darnell, J. E. *Biochemistry* **17**, 2776–2783 (1978).
4. Fedoroff, N., Wellauer, P. K. & Wall, R. *Cell* **10**, 597–610 (1977).
5. Fedoroff, N. & Wall, T. R. in *Molecular Mechanisms in the Control of Gene Expression*, 379–384 (Academic, New York, 1976).
6. Smith, M. J., Hough, B. R., Chamberlin, M. E. & Davidson, E. H. *J. molec. Biol.* **85**, 103–126 (1974).
7. Holmes, D. S. & Bonner, J. *Proc. natn. Acad. Sci. U.S.A.* **71**, 1108–1112 (1974).
8. Costantini, F. D., Scheller, R. H., Britten, R. J. & Davidson, E. H. *Cell* **15**, 173–187 (1978).
9. Costantini, F. D., Britten, R. J. & Davidson, E. H. *Nature* **287**, 111–117 (1980).
10. Anderson, D. M. *et al. J. molec. Biol.* **155**, 281–309 (1982).
11. Spohr, G., Reith, W. & Sures, I. *J. molec. Biol.* **151**, 573–592 (1981).
12. Davidson, E. H., Hough, B. R., Klein, W. H. & Britten, R. J. *Cell* **4**, 217–238 (1975).
13. Klein, W. H., Murphy, W., Attardi, G., Britten, R. J. & Davidson, E. H. *Proc. natn. Acad. Sci. U.S.A.* **71**, 1785–1789 (1974).
14. Campo, M. S. & Bishop, J. O. *J. molec. Biol.* **90**, 649–663 (1974).
15. Dina, D., Meza, I. & Crippa, M. *Nature* **248**, 486–490 (1974).
16. Kindle, K. L. & Firtel, R. A. *Nucleic Acids Res.* **6**, 2403–2422 (1979).
17. Kimmel, A. R. & Firtel, R. A. *Cell* **16**, 787–796 (1979).
18. Zuker, C. & Lodish, H. F. *Proc. natn. Acad. Sci. U.S.A.* **78**, 5386–5390 (1981).
19. Calabretta, B., Robberson, D. L., Maizel, A. L. & Saunders, G. F. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6003–6007 (1981).
20. Houck, C. M., Rinehart, F. P. & Schmid, C. W. *J. molec. Biol.* **132**, 289–306 (1979).
21. Jelinek, W. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 1398–1402 (1980).
22. Weiner, A. M. *Cell* **22**, 209–218 (1980).
23. Haynes, S. R. & Jelinek, W. R. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6130–6134 (1981).
24. Duncan, C. H., Jagadeeswaran, P., Wang, R. R. C. & Weissman, S. M. *Gene* **13**, 185–196 (1981).
25. Elder, J. T., Pan, J., Duncan, C. H. & Weissman, S. M. *Nucleic Acids Res.* **9**, 1171–1188 (1981).
26. Pearson, W. R., Wu, J. -R. & Bonner, J. *Biochemistry* **17**, 51–59 (1978).
27. Davidson, E. H. & Britten, R. J. *Q. Rev. Biol.* **48**, 565–613 (1973).
28. Crampton, J. M., Davies, K. E. & Knapp, T. F. *Nucleic Acids Res.* **9**, 3821–3834 (1981).
29. Posakony, J. W., Scheller, R. H., Anderson, D. M., Britten, R. J. & Davidson, E. H. *J. molec. Biol.* **149**, 41–67 (1981).
30. Scheller, R. H., Costantini, F. D., Kozlowski, M. R., Britten, R. J. & Davidson, E. H. *Cell* **15**, 189–203 (1978).
31. Ernst, S. G., Britten, R. J. & Davidson, E. H. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2209–2212 (1979).
32. Kleene, K. C. & Humphreys, T. *Cell* **12**, 143–155 (1977).
33. Wold, B. J., Klein, W. H., Hough-Evans, B. R., Britten, R. J. & Davidson, E. H. *Cell* **14**, 941–950 (1978).
34. Potter, S. S., Brorein, W. J., Dunsmuir, P. & Rubin, G. M. *Cell* **17**, 415–427 (1979).
35. Strobel, E., Dunsmuir, P. & Rubin, G. M. *Cell* **17**, 429–437 (1979).
36. Dunsmuir, P., Brorein, W. J., Simon, M. A. & Rubin, G. M. *Cell* **21**, 575–579 (1980).
37. Levis, R., Dunsmuir, P. & Rubin, G. M. *Cell* **21**, 581–588 (1980).
38. Flavell, A. J., Ruby, S. W., Toole, J. J., Roberts, B. E. & Rubin, G. M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 7107–7111 (1980).
39. Scherer, G., Telford, J., Baldari, C. & Pirotta, V. *Devl Biol.* **86**, 438–447 (1981).
40. Levine, M., Garen, A., Lepesant, J. -A. & Lepesant-Kejzlarova, J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2417–2421 (1981).
41. Britten, R. J. & Davidson, E. H. *Science* **165**, 349–358 (1969).
42. Errede, B., Cardillo, T. S. & Sherman, F. *Cell* **25**, 427–436 (1980).
43. Orgel, L. & Crick, F. H. C. *Nature* **284**, 604–607 (1980).
44. Doolittle, W. F. & Sapienza, C. *Nature* **284**, 601–603 (1980).
45. Britten, R. J. & Davidson, E. H. *Q. Rev. Biol.* **46**, 111–138 (1971).
46. Moore, G. P., Scheller, R. H., Davidson, E. H. & Britten, R. J. *Cell* **15**, 649–660 (1978).
47. Calabretta, B., Robberson, D. L., Barrera-Saldana, H. A., Lambrout, T. P. & Saunders, G. F. *Nature* **296**, 219–225 (1982).
48. Krolewski, J. J., Bertelsen, A. H., Humayun, M. Z. & Rush, M. G. *J. molec. Biol.* (in the press).
49. Flavell, A. J. & Ish-Horowicz, D. *Nature* **292**, 591–595 (1981).
50. Ryffel, G. U., Muellener, D. B., Wyler, T., Wahli, W. & Weber, R. *Nature* **291**, 429–431 (1981).
51. Page, G. S., Smith, S. & Goodman, H. M. *Nucleic Acids Res.* **9**, 2087–2104 (1981).
52. Cochet, M., Gannon, F., Hen, R., Maroteaux, L., Perrin, F. & Chambon, P. *Nature* **282**, 567–574 (1979).
53. Davidson, E. H. & Britten, R. J. *Science* **204**, 1052–1059 (1979).
54. Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. & Steitz, J. A. *Nature* **283**, 220–224 (1979).
55. Calvet, J. P., Meyer, L. M. & Pederson, T. *Science* (in the press).

ARTICLES

Climatic significance of the hydrogen isotope ratios in tree cellulose

Crayton J. Yapp* & Samuel Epstein

Division of Geological and Planetary Sciences, California Institue of Technology, Pasadena, California 91125, USA

A distinct relationship exists between the δD values of cellulose carbon bound hydrogen in trees and average annual temperature for a variety of trees from a wide range over North America. The slope of the δD/T relationship is 5.8‰ °C^{-1}. Samples of annual precipitation covering much the same geographical range as the trees exhibit a comparable temperature coefficient of 5.6‰ °C^{-1}. Such growth-site conditions as poor drainage and/or low relative humidity seem to perturb the spatial δD/T relationship. However, our data indicate that suitable tree growth sites are more the rule than the exception.

A GROWING body of evidence suggests that variations in the D/H ratios of the carbon-bound hydrogen of tree cellulose primarily reflect variations in the D/H ratio of associated meteoric waters as manifested in the isotopic composition of the leaf water[1-3]. Because there is a generally recognized relationship between climatic temperature and the stable hydrogen and oxygen isotope composition of meteoric water[4-6], there is an expectation that the δD variations of tree cellulose C–H hydrogen will reflect variations in climate

$$\delta D = \left(\frac{R_{\text{sample}}}{R_{\text{standard}}} - 1\right) \times 1{,}000, \text{ where } R = D/H \text{ ratio.}$$

The standard is SMOW (standard mean ocean water as defined in ref. 7). Such a climatic 'signal' has already been identified for tree cellulose $^{18}O/^{16}O$ ratios[8-10].

The few studies that have been done on the climatic significance of tree cellulose C–H hydrogen δD variations[11-14] have not been completely satisfactory, largely because of experimental methodology[1,15,16]. With the development of techniques that yield highly reproducible δD values (refs 3, 17), the opportunity has arisen for further investigations into the climatic significance of the D/H ratios of tree cellulose carbon-bound hydrogen (note that our analytical method has a δD precision of ±2‰).

When considering the climatic significance of δD variations in trees, it is important to keep in mind that there are two scales of climatic change which are of interest—spatial and temporal. The δD variations in trees offer the possibility that both spatial and temporal climatic signals can be discerned, because trees are geographically widespread and also contain an internal chronology which may isotopically record local climatic change. This article is concerned with the climatic significance of tree δD values on the spatial scale. The question of the climatic significance of δD variations within the chronologies of individual trees will be considered elsewhere.

Results and discussion

Twenty-five trees from 22 sites throughout North America were obtained and analysed for their cellulose C–H hydrogen δD value by the renitration method[3]. The outermost four or five rings made up the samples which were analysed. Because of the different times at which the various trees were felled, the total interval spanned by the samples is AD 1961–75. Sample descriptions can be found in ref. 3. The locations of the trees as well as the North American sites of the International Atomic Energy Agency (IAEA) precipitation stations are plotted in Fig. 1. It can be seen in Fig. 1 that the tree samples cover much the same range of geography and climate as the IAEA stations. Thus, if the δD variations of the tree cellulose C–H hydrogen reflect the δD values of local precipitation, the two should have similar relationships to spatial variations in climatic temperature.

The average annual temperature and precipitation δD values of the IAEA sites of Fig. 1 are plotted in Fig. 2. The relationship between these two variables is readily apparent, and a linear regression of these data yields the expression:

$$\delta D = 5.6T\ (°C) - 120 \tag{1}$$

with a correlation coefficient of 0.90. The value of 5.6 for the slope of this equation is the same as that found by Dansgaard[4] for maritime and Arctic stations, although the intercept of the North American data of Fig. 2 is more negative by about 20‰.

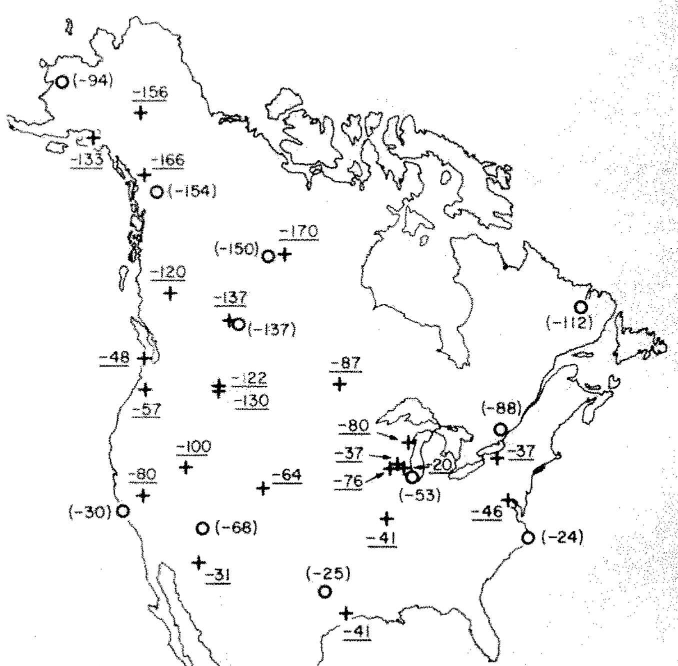


Fig. 1 The geographical distribution of the δD values of tree samples analysed for this work. In addition, the δD values of annual precipitation at the 11 North America IAEA stations are depicted. +, Tree sites; ○, IAEA stations. δD values of the trees are underlined, while the δD values of the precipitation at the IAEA stations are shown in parentheses. Note that there is a general decrease in δD values of both the trees and the precipitation from coastal to inland sites and from south to north.

* Present address: Department of Geology, University of New Mexico, Albuquerque, New Mexico 87131, USA.

0028-0836/82/250636—04$01.00

The δD values measured for the trees and the average annual temperatures of nearby meteorological stations are listed in Table 1 and plotted in Fig. 3 which shows that there is a relationship between the δD values and annual temperature although some scatter is also evident. The five circled points in Fig. 3 seem to lie significantly above the general 'trend line' of the data and may be unusual from the standpoint of the conditions in their environments of growth. For example, UBLM-2 is a willow tree which grew in a marsh in Madison, Wisconsin. The waters in the soil in which this tree grew were analysed and found to have a δD value of −32‰. This value is considerably more positive than the average annual precipitation in the area, the latter of which has a δD value of about −60‰. Consequently, either the meteoric water incorporated by the tree had undergone significant evaporation as it slowly infiltrated into the poorly drained soil, or the water represented predominantly summer precipitation which is commonly more positive than the average annual precipitation[18]. In either case, the δD value recorded by the tree would not be representative of average annual precipitation and therefore would probably not reflect average annual temperature.

Samples MNY-GA-2, MNY-RM-5 and MNY-BO-7 (encircled in Fig. 3) grew in a marsh in New York. The waters utilized by these plants were not available for analysis, but conditions similar to those noted for the UBLM-2 site were possibly present at the site of these three trees. They may then have incorporated waters with δD values which were more positive than the average annual δD values of precipitation in the area and therefore the trees would not have δD values representative of the local average annual temperature.

The fifth point encircled in Fig. 3 is AW-BO-1 a bur oak which grew in a pasture in southeastern Wisconsin amidst a small stand of oaks. The analysed δD value of water from a nearby stream was −62‰. This water value is similar to δD values measured for water from the Wisconsin River and from a river near Oconto, Wisconsin. In the latter two cases, trees (WR-2, OM-2) which grew in soils immediately adjacent to these waters had δD values in the range −76 to −80‰. Consequently, if AW-BO-1 had incorporated waters with a δD value of −62‰, it would have been expected to have a δD value around −75‰. In fact AW-BO-1 has a δD value of −23‰. Unless there is some unknown biochemical or physiological factor present in bur oaks which is causing this comparatively positive δD value, it appears that this tree incorporated waters with δD values more positive than the average annual value. As with UBLM-2, this might be the case if the root system were sufficiently near the surface to cause the tree to use only summer precipitation. Alternatively, the infiltration rate of water into the soil may be slow enough in the environs of the tree to allow significant increase in δD values of precipitation due to evaporation while on the soil surface (see ref. 19).

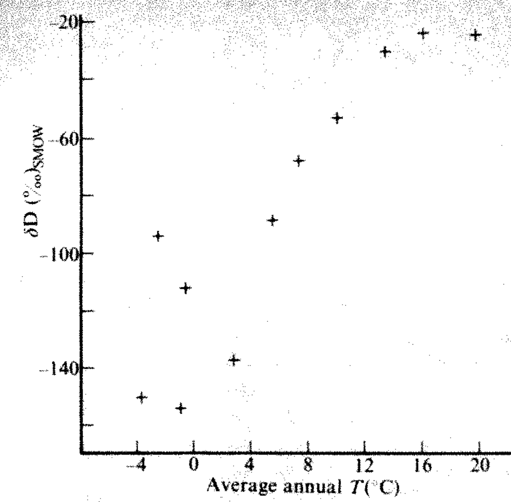


Fig. 2 A plot of the δD values of annual precipitation against annual temperature for 11 available North American IAEA stations. The linear regression of these data yields the equation: $\delta D = 5.6T\ (°C) - 120$. The correlation coefficient is 0.90.

As discussed in ref. 19, this soil surface evaporation could cause a greater increase in δD values of the water than would occur from transpiration alone. Thus, significant above-ground evaporation of precipitation might cause tree δD values to be more positive than otherwise expected. The relationships of these more positive tree δD values to annual temperature in a spatial $\delta D/T$ plot would probably differ from that of other sites at which no above-ground evaporation of precipitation had occurred.

In short, the circled points in Fig. 3 are interpreted to be representative of sites at which trees incorporated soil water that was more positive than the δD value of average annual

Table 1 δD of cellulose nitrate and approximate average annual temperature in region of growth site—all sites are North American

Sample no.	Name	Location	Average annual T (°C)	Cellulose nitrate δD
OPW-DF-1	Sitka spruce	Olympic Peninsula, Washington	11.3	−48
Col-DF-1	Douglas fir	Jefferson Co., Colorado	10.5	−64
MNY-GA-2	Green Ash	Montezuma National Wildlife	8.2	−33
MNY-BO-7	Bur Oak	Refuge,	8.2	−34
MNY-RM-5	Red Maple	New York	8.2	−45
RE-CO-2	Chestnut oak	Reston, Virginia	14.0	−46
BCT-12	Pine	Seeley Lake, British Columbia	3.4	−120
MO-O-2	Oak	Owensville, Missouri	12.6	−41
AW-BO-1	Bur Oak	Albion, Wisconsin	7.0	−20
FCA-WS-1	White Spruce	Fox Creek, Alberta	0.9	−137
SAS-JP-1	Jack Pine	Porter Lake, Northwest Territories	−3.5	−170
FA-As-1	Aspen	Fairbanks, Alaska	−3.3	−156
KSA-As-1a	Aspen	Anchorage, Alaska	2.1	−133
NT-1	Aspen	Brown Creek, Ruby Mts, Nevada	7.4	−100
MT-1	Birch	Flathead Lake, Montana	5.7	−130
MT-2	Maple	Flathead Lake, Montana	5.7	−122
UAZ-PP-1	Ponderosa pine	Catalina Mts, Arizona	19.3	−31
PO-DF-1	Douglas fir	Portland, Oregon	12.6	−57
H-2	Oak	Houston, Texas	20.2	−41
WM-B-1	Birch	Winnipeg, Manitoba	2.5	−87
WR-2	Maple	Spring Green, Wisconsin	7.0	−76
OM-2	White Birch	Oconto, Wisconsin	7.0	−80
UBLM-2	Willow	Madison, Wisconsin	7.0	−37
Yuk-W-1	White Spruce	Kluane Lake, Yukon Territory	−0.7	−166
BrPn	Bristlecone pine	White Mts, California	0.7	−80

precipitation. Thus, these samples suggest that site conditions may be very important in choosing trees for possible interpretation of geographical climatic variation. The scatter in Fig. 3 not accounted for by the circled points will be discussed below.

If the circled points in Fig. 3 are eliminated from consideration in the spatial $\delta D/T$ plot, the balance of the data produces an array of points for which a linear regression yields the expression:

$$\delta D = 5.8T\ (^{\circ}C) - 134 \quad (2)$$

The linear correlation coefficient is 0.88. The slope of this curve is very similar in magnitude to the slope of 5.6 obtained for the regression of North American IAEA precipitation data (Fig. 2). Because the tree sites and IAEA stations extend over much the same range of geography and climate (Fig. 1), it might be expected that a regression of the two sets of δD data against temperature would produce a similar slope, providing that the δD values of the tree samples used in the comparison reflect the δD values of annual precipitation at these sites.

Seven of the tree samples, OPW-DF-1, Col-DF-1a, RE-CO-2, BCT-12, MO-O-2, FCA-WS-1, and SAS-JP-1, included in the regression of the data of Fig. 3 have chronologies that extend back to at least AD 1931. They range in location from as far south as St Louis, Missouri, and Reston, Virginia, to as far north as Porter Lake, Northwest Territories.

In addition, these seven trees are located near meteorological stations whose instrumental temperature records also date from at least 1931. This suggests the possibility of plotting spatial tree δD variations against annual temperature for 5-yr intervals from 1931 to 1970. Such plots could determine whether or not there have been any important changes over this interval in the character of the $\delta D/T$ relationship as defined by the seven trees (see Fig. 4).

Because far fewer sites are included in Fig. 4 than in Fig. 3, it is statistically possible that the linear regressions of these data will not produce the same slope as that obtained from the regression of the data of Fig. 3. Thus, it does not seem to be possible to attach physical significance to any differences in slope between the large data group (Fig. 3) and its subset (Fig. 4). However, changes of slope between different 5-yr intervals in Fig. 4 may have a physical interpretation.

The equations for the linear regressions of the different 5-yr intervals are found in Fig. 4. The analytical error of ±2‰ for the measurement of δD values imposes an error on the regression slopes of Fig. 4 of about ±0.3‰ $^{\circ}C^{-1}$. Thus, with the exception of 1941–45 all the 5-yr intervals exhibit within experimental error the same spatial δD temperature coefficients. The large 1941–45 temperature coefficient of 8.6‰ $^{\circ}C^{-1}$ seems to be a consequence of something other than normal analytical error. However, at present, we cannot explain it.

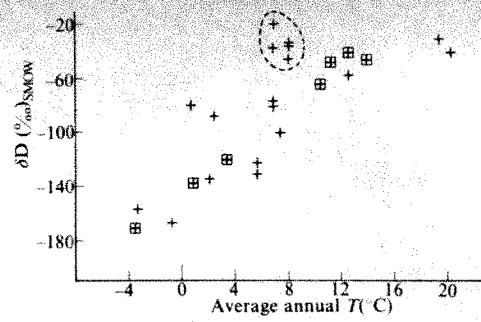


Fig. 3 A plot of the δD values of trees from a wide range over North America against the average annual temperature. The tree samples represent various intervals within the span 1961–75. The five points enclosed in the dashed circle represent trees which may have grown in conditions that could obscure a climatic interpretation of their δD values. The remaining 20 points define a trend with the following linear regression equation: $\delta D = 5.8T\ (^{\circ}C) - 134$. The correlation coefficient is 0.88. ⊞, Trees in Fig. 4.

Notwithstanding the possibility that the spatial 1941–45 $\delta D/T$ slope is truly different from the others, the fundamental observation to be drawn from Fig. 4 is that over this entire 40-yr period, an impressively consistent pattern of declining tree δD values with declining spatial temperature has been maintained. This has occurred in spite of the differences in tree species and different relative ages of the trees. Thus, it enhances the conclusion that the δD values of trees can indeed be indicators of spatial climatic differences.

Some of the possible causes of the scatter in the 20 'well-behaved' data points of Fig. 3 should now be considered. It is possible, for example, that trees growing in regions of low humidity may exhibit comparatively positive δD values due to a high degree of deuterium enrichment in the leaf water during transpiration. This would tend to shift the cellulose carbon-bound hydrogen δD value above the general trend of the data of Fig. 3, even if the tree is using soil water whose isotopic composition is equal to that of the average annual precipitation[3]. The relatively positive position of the BrPn point in Fig. 3 may be explained by this effect, as the average growing season

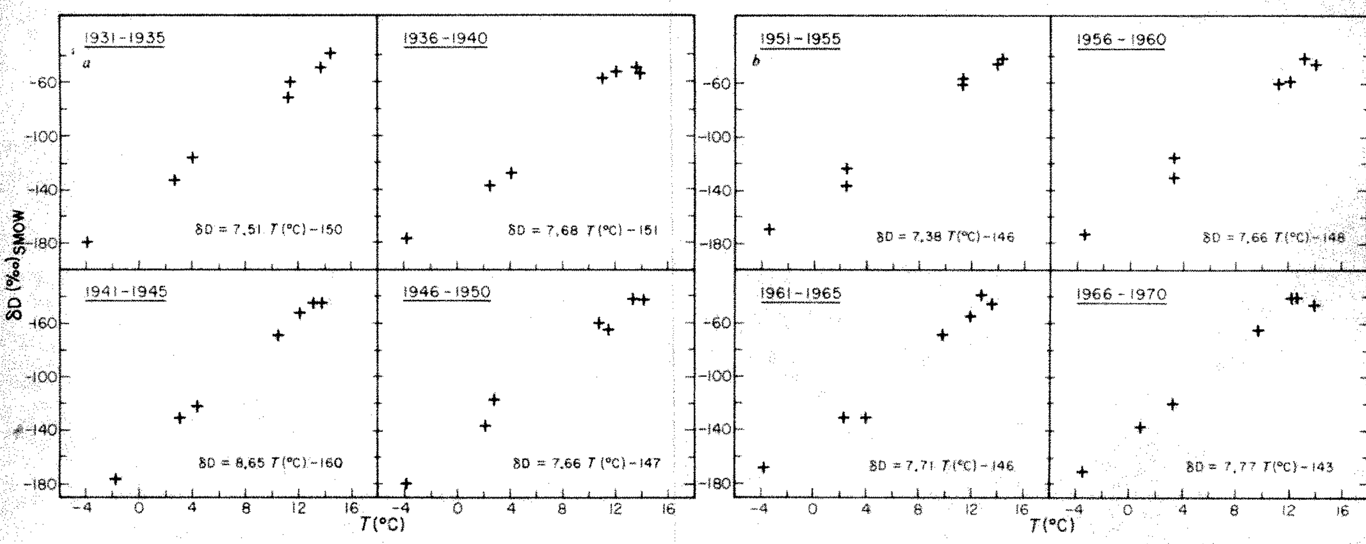


Fig. 4 *a*, Comparisons of the δD values of seven trees from a wide range over North America (OPW-DF-1, Col-DF-1a, RE-CO-2, BCT-12, MO-O-2, FCA-WS-1, and SAS-JP-1) with average annual temperature. The comparisons are made for different corresponding 5-yr intervals. The equations of the linear regressions of the data are shown within the plots of each of the 5-yr intervals. The pattern of declining δD values with declining temperature is remarkably consistent for the trees and the intervals depicted. *b*, Same trees as in *a* except that four different 5-yr intervals are compared. Again, the overall pattern of declining δD with declining temperature is maintained throughout these intervals in spite of the differences in species and relative ages of the trees.

relative humidity in the region of BrPn is only about 45% (ref. 3). Interestingly, it may eventually be possible to correct for the deuterium enrichment due to transpiration by analysing both hydrogen and oxygen isotope ratios in cellulose (see ref. 10).

A second source of scatter may arise from the fact that the surface temperatures used for comparison with the tree δD values are unweighted average annual temperatures. If the trees are incorporating annual precipitation, the δD value of this precipitation will probably be a weighted average of varying proportions of the various seasonal precipitations. Consequently, the unweighted temperature and weighted δD values may not always be uniformly related. Furthermore, the compared temperatures are those of meteorological stations which are not located at the exact site of tree growth. Thus, the pattern of geographical temperature variation of these stations may not be exactly analogous to that of the growth sites.

Biochemical or physiological differences between these 20 different 'well-behaved' trees could produce some, as yet, unrecognized differences in the net hydrogen isotope fractionation between leaf water and cellulose C–H hydrogen. Such differences might contribute to the scatter in Fig. 3.

There may also be trees among the 20 'well-behaved' samples that have not incorporated average annual precipitation. Instead, some trees may have utilized relatively more water from a single season. When plotted against annual temperature, the δD values of these trees might produce some scatter such as appears in Fig. 3.

Finally, there is likely to be scatter in the primary precipitation δD/temperature relationship (see Fig. 2). This scatter would be manifested in Fig. 3, if the δD values of the trees reflected the δD values of the local meteoric waters.

In spite of these possible sources of scatter in the relationship between tree δD values and annual temperature, Fig. 3 shows that such a relationship does exist. Furthermore, for the 20 'well-behaved' samples of Fig. 3, the relationship is quite close to that which might be expected if the δD variations of the trees mimic those of the local annual precipitation.

Thus, trees from the proper sites seem to record δD values that correspond to geographical climatic differences. Note that the samples analysed for this work were collected by individuals who had little specific instruction regarding the relative merits of locat tree sites. Thus, local site selection was largely random. From information supplied by individuals who collected the samples (see ref. 3), it seems that the 'well-behaved' sites are likely to be those in which there is an absence of stagnant soil, ground or surface water[19].

Conclusion

A comparison of the geographical variation of the δD values of trees with the associated average annual temperature reveals an overall correlation for many sites and species of trees. The spatial linear temperature coefficient obtained for δD values from 20 widely separated tree samples in North America is 5.8‰ °C^{-1}. This compares with a temperature coefficient of 5.6‰ °C^{-1} obtained from annual precipitation δD values at 11 North American IAEA sites. The similarity of these two coefficients is consistent with the idea that the δD variations of the trees reflect the δD variations of the local precipitation. The latter conclusion is supported by the spatial distribution of the tree δD values (Fig. 1). These δD values decrease from the coast to the interior and from south to north. This pattern of variation is analogous to that observed for meteoric waters over North America (Fig. 1).

Site conditions seem to be important in determining the δD values recorded by trees, and some site conditions can produce scatter in plots of annual temperatures against tree δD values on a spatial scale. However, the trees sampled for this work indicate that favourable growth sites are common.

The geographical temperature–tree δD correlation for seven widely separated North American trees shows a remarkable consistency over the interval 1931–70. This consistency is maintained in spite of the differences in both species and relative ages of the seven tree samples involved.

We thank everyone who provided samples for this work (see ref. 3 for names). NSF grant ATM80-18830 supported this research. This is contribution no. 3691 of the Division of Geological and Planetary Sciences, California Institute of Technology.

Received 31 December 1981; accepted 18 March 1982.

1. Epstein, S., Yapp, C. J. & Hall, J. H. *Earth planet. Sci. Lett.* **30,** 241 (1976).
2. DeNiro, M. J. & Epstein, S. *Science* **204,** 51 (1979).
3. Yapp, C. J. & Epstein, S. *Geochim cosmochim Acta* (submitted).
4. Dansgaard, W. *Tellus* **16,** 436 (1964).
5. Friedman, I., Redfield, A. C., Schoen, B. & Harris, J. *Rev. Geophys.* **2,** 177 (1964).
6. Siegenthaler, U. & Oeschger, H. *Nature* **285,** 314 (1980).
7. Craig, H. *Science* **133,** 1833 (1961).
8. Gray, J. & Thompson, P. *Nature* **262,** 481 (1976).
9. Burk, R. L. & Stuiver, M. *Science* **211,** 1417 (1981).
10. Epstein, S., Thompson, P. & Yapp, C. J. *Science* **198,** 1209 (1977).
11. Schiegl, W. E. *Nature* **251,** 582 (1974).
12. Libby, L. M. & Pandolfi, L. J. *Proc. natn. Acad. Sci. U.S.A.* **71,** 2482 (1974).
13. Wilson, A. T. & Grinsted, M. J. *Nature* **257,** 387 (1975).
14. Epstein, S. & Yapp, C. J. *Earth planet. Sci. Lett.* **30,** 252 (1976).
15. Epstein, S. & Yapp, C. J. *Nature* **266,** 477 (1977).
16. Epstein, S. *Earth planet. Sci. Lett.* **39,** 303 (1978).
17. DeNiro, M. J. *Earth planet. Sci. Lett.* **54,** 177 (1981).
18. *Technical Report Series* Nos. 96, 117, 147 (IAEA, Vienna 1969, 1970, 1973).
19. Yapp, C. J. thesis, California Inst. Technol. (1980).

β Decay and the origins of biological chirality: experimental results

D. W. Gidley, A. Rich & J. Van House

Physics Department, University of Michigan, Ann Arbor, Michigan 48109, USA

P. W. Zitzewitz

Department of Natural Sciences, University of Michigan-Dearborn, Dearborn, Michigan, 48128, USA

A spin-polarized low-energy positron beam has been used to set limits on asymmetric positronium formation in optically active molecules. No asymmetry was found at the 7×10^{-4} level in cystine and tryptophan, but a possible effect of $(31\pm7)\times10^{-4}$ was found in leucine. A quantitative connection is made with the origin of biological optical activity.

THE amino acids and sugars on which terrestrial life is based show maximal optical activity, that is, with rare exceptions, they are composed of D-sugars in RNA and DNA and L-amino acids in proteins. This observation poses the questions: (1) Why should life be based on optically active substances? (2) Are there any causal mechanisms that would lead us to expect the selection of the L-amino acids in terrestrial organisms and is this choice expected to be the one statistically preferred when all possible biospheres are considered? Question (1) has been discussed extensively[1,2] and there is widespread agreement that life should naturally select an ordered system based on optically pure substances. However, question (2), still unanswered, is one of the important problems in chemical and biological evolution.

In the absence of any causal mechanisms, the presently observed complete optical purity would be the result of a random fluctuation in the presumed virtually racemic isomeric balance of the primordial Earth, followed by chemical and biological amplification. The probability of selecting a particular isomer would then be 50%. On the other hand, causal mechanisms can systematically produce an isomeric excess which, if large enough to compete with random fluctuations, may strongly bias the odds on which isomer will eventually dominate. To generalize our treatment of question (2) to all types of biospheres it is necessary to distinguish two types of causal mechanisms: local and universal. A local causal mechanism relies on some local asymmetry to produce a bias in the odds on which isomer dominates in that particular biosystem. By contrast, a universal causal mechanism produces a systematic bias in the odds in favour of one isomer dominating throughout all biosystems in which it is effective. All such mechanisms must be related to the weak interaction as this is the only interaction that universally violates parity conservation. No reproducible quantitative results are yet available for either the random or causal mechanisms. The description, theoretical analysis, and preliminary experimental results of a new method to investigate the most plausible universal causal mechanism, preferential radiolysis by electrons emitted in the β decay of radionuclides, is the subject of this and the accompanying article[3]. For completeness, we also present a brief discussion of the leading local causal mechanism, preferential photolysis or catalysis by circularly polarized sunlight.

Analysis of two causal mechanisms

Preferential photolysis by circularly polarized light: Asymmetric effects have been observed in the interaction between amino acid isomers and circularly polarized light[4]. In addition, a circular polarization (CP) of sunlight of 0.1% has been observed at dawn and dusk in an IR band (780–800 nm)[5]. However, a corresponding CP in the UV, needed for asymmetric photolysis, was not found, and thus it must be <0.01%. In addition, because CP is a result of parity conserving electromagnetic interactions in the atmosphere, it can only be a local causal mechanism, that is, it must be identically zero when averaged over one Earth rotation and over the Earth's surface, unless non-uniformities exist in time (morning compared with evening) and with respect to the Northern and Southern Hemispheres.

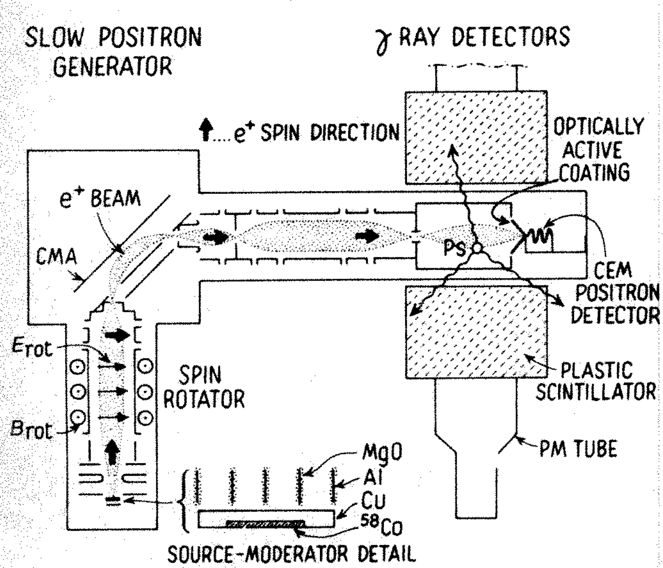


Fig. 1 The experimental apparatus used to measure asymmetries in triplet positronium formation. The beam consists of 3×10^3 positrons s^{-1} and initial helicity $h_0(e^+)=0.21\pm0.02$ (ref. 13). The Wien filter spin rotator (crossed electric and magnetic fields) allows rotation of the average spin direction ($\langle\hat{\mathbf{s}}_i\rangle$) of the beam with minimal effect on the average direction of the beam's momentum ($\langle\hat{\mathbf{p}}_i\rangle$). Thus $h_0(e^+)$ may be continuously varied from +0.21 to −0.21.

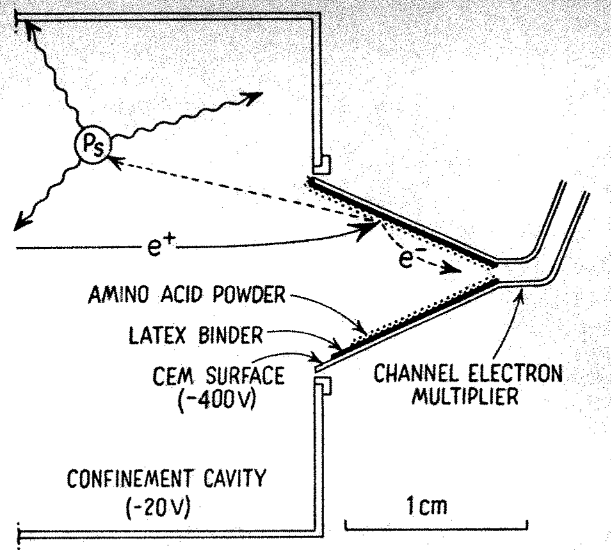


Fig. 2 Detail of the slow positron target region. A 0.5-mm thick layer of powder of a given sample was pressed on a binding layer of latex on the CEM cone. Measurements of Ps formation on the bare cone, the latex binder, and the amino acids confirmed that Ps formation occurred only in the amino acid layer. Secondary electrons are collected into the channel of the CEM. The triplet Ps formed in the amino acid powder lives a sufficient time to escape from the powder into the confinement cavity and annihilate into three γ rays with approximately the vacuum lifetime of 140 ns.

Preferential radiolysis related to electron helicity in β decay: The radiolysing β decay electrons, on emission from the nucleus, possess a handedness or helicity (h), that is—there is a correlation between the direction of the spin angular momentum and the linear momentum. This correlation is expressed by the relation $h=\langle h_i\rangle=\langle\hat{\mathbf{s}}_i\cdot\hat{\mathbf{p}}_i\rangle$ where, for the ith particle, $\hat{\mathbf{s}}_i$ is the Pauli spin matrix, $\hat{\mathbf{p}}_i$ the unit momentum vector ($\hat{\mathbf{p}}_i=\vec{\mathbf{p}}_i/|\vec{\mathbf{p}}_i|$), h_i is the helicity operator, and the angular brackets represent an average over all particles in the ensemble. Further, for the present discussion $\langle\hat{\mathbf{s}}_i\cdot\hat{\mathbf{p}}_i\rangle=\langle\hat{\mathbf{s}}_i\rangle\cdot\langle\hat{\mathbf{p}}_i\rangle$, that is the positron spin and momenta are decoupled. If, for example, all spins and all momenta in a beam are parallel $|\langle\hat{\mathbf{s}}_i\rangle|=1$ and $|\langle\hat{\mathbf{p}}_i\rangle|=1$. Preferential radiolysis by β decay electrons represents what we have defined as a universal causal mechanism and the problem is to determine to what degree this handed radiation could asymmetrically radiolyse racemic mixtures of amino acids on the primordial Earth.

Many experiments have searched for helicity induced preferential radiolysis[6–8] but no reproducible effect has been demonstrated. We feel that this is because none of the experiments approached the level of sensitivity necessary to observe the small asymmetric effects which are now predicted to occur[3]. These effects arise from the distortion in the electronic wavefunctions in optically active molecules. The distortion induced by coupling between spin and orbital motion in the bound electrons produces a helicity per unit volume or helicity density—denoted $h(\vec{\mathbf{r}})$. From considerations of parity conservation (mirror symmetry) $h(\vec{\mathbf{r}})$ is zero for symmetric spin-unpolarized molecules but $h(\vec{\mathbf{r}})$ can be non-zero for spin-unpolarized dissymmetric molecules.

The spin-dependent exchange interaction between the molecular electrons and the incident β decay electrons, when the coupling between $h(e^-)$, the helicity of the incident electrons, and $h(\vec{\mathbf{r}})$ is included, produces an asymmetry, A_R, in the respective rates of radiolysis of L and D isomers $R(\text{L})$, $R(\text{D})$. On either reversal of $h(e^-)$ ($h(e^-)\to -h(e^-)$) or interchange of the L and D isomers this asymmetry may be written as

$$A_R\equiv\frac{R^+(\text{L})-R^-(\text{L})}{R^+(\text{L})+R^-(\text{L})}=\frac{R^+(\text{L})-R^+(\text{D})}{R^+(\text{L})+R^+(\text{D})}$$
$$=|h(e^-)|H_R(E,Z) \qquad (1)$$

Here (+, −) refers to $h(e^-)>0$ or $h(e^-)<0$ and from parity conservation $R^{\pm}(\text{D}) = R^{\mp}(\text{L})$ so that, for example, A_R may be written as:

$$A_R = [R^-(\text{D}) - R^+(\text{D})]/[R^-(\text{D}) + R^+(\text{D})]$$

H_R, an effect of the molecular helicity density $h(\mathbf{\bar{r}})$, is essentially an average of $h(\mathbf{\bar{r}})$ weighted by the Coulomb interaction between the projectile (β) electron and the target (molecular) electrons. For projectile energies E_p of the order of 100–500 keV, H_R is given by the relation

$$H_R \simeq \eta'_R (\alpha Z)^2 \frac{1}{2E_p \ln(2E_p)}$$

where α is the fine structure constant ($\alpha^{-1} \sim 137$), Z the atomic number of the heaviest atom in the asymmetric environment of the molecule and η'_R is a molecular asymmetry factor which takes into account the effect of molecular structure on H_R. See the accompanying article for a complete discussion of H_R. We note here that at $E_p \sim 100$ keV and $Z = 6$, H_R lies in the range $10^{-11} < H_R < 10^{-10}$. The uncertainty in H_R is primarily a result of the uncertainties of the molecular wave functions used in its calculation. Thus even if we take $h(e^-) = 1$ and assume that it does not decrease as the β slows down, (in fact $h(e^-) \to 0$ as the β slows down due to velocity dispersion caused by scattering[9]), the largest resulting value of A_R ($A_R < 10^{-10}$) would still be much too small to be observed in a direct radiolysis experiment. Current techniques limit such experiments to detection of $A_R > 10^{-3}$.

Larger effects, though less directly related to $h(\mathbf{\bar{r}})$, may in fact be obtained by performing scattering experiments using unpolarized electron beams whose energy is of the order of the molecular ionization potential (I). A phenomenological analysis of this possibility has been presented[10] and quantitative estimates of the effects to be expected (an induced helicity between 10^{-3} and 10^{-5} in an initially unpolarized electron beam) have been made[25]. These estimates show that observation of the helicity is not possible using available techniques, a result corroborated by the preliminary experiments sensitive to an induced helicity of order 10^{-2} which have not seen an effect[11].

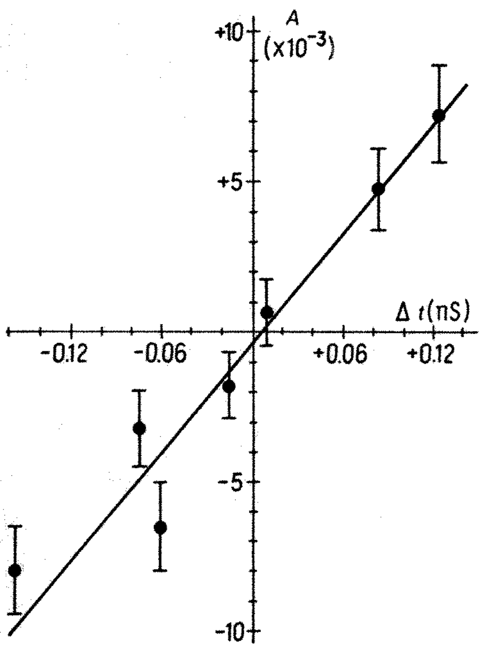


Fig. 3 The observed asymmetry in triplet Ps formation on DL cystine plotted against $\Delta\tau$, the observed time difference between the prompt peaks after helicity reversal. Values of r^+ and r^- based on $\sim 16 \times 10^6$ prompt time events and 10^6 delayed events were used to calculate A_{Ps}, according to equation (4). The uncertainties in A_{Ps} for a 12-h sequence were typically of order 10^{-3} based on Poisson ($\sqrt{N}$) statistics for prompt and delayed counts. The straight line is the linear least squares fit.

Table 1 The residual asymmetry, $A_{Ps}(0)$, and the χ^2 per degree of freedom obtained from the least squares fitting procedure

Substance		D	L	DL
Cystine	$A_{Ps}(0)$ ($\times 10^{-4}$)	−10.8 (5.1)	−3.6 (5.0)	−4.2 (4.9)
	χ^2/N	4.8/5	7.4/7	6.2/7
Leucine	$A_{Ps}(0)$ ($\times 10^{-4}$)	−15.1(7.3)	+4.0 (6.6)	−3.3 (3.2)
	χ^2/N	4.4/5	7.3/8	7.1/6
Tryptophan	$A_{Ps}(0)$ ($\times 10^{-4}$)	−2.7 (4.6)	−1.3 (5.5)	−10.9 (3.1)
	χ^2/N	0.7/3	3.9/4	7.0/7

In our experiment we employ a method that is much more sensitive to $h(\mathbf{\bar{r}})$ than the radiolysis experiments. A beam of 200–400 eV positrons with a net helicity, $h(e^+)$, is directed into an amino acid target. After slowing down to ~10 eV, 80% of the positrons annihilate directly with an electron. The remaining 20% capture an electron to form positronium (Ps), the hydrogen-like bound state of the two particles. The Ps is formed in both the triplet (140 ns lifetime) and singlet (0.1 ns lifetime) spin states and the experiment is designed to detect only the triplet state (based on its long lifetime). Because of the existence of $h(\mathbf{\bar{r}})$ in the molecule, the fraction of this state formed for a given isomer $f(\text{L})$ or $f(\text{D})$ depends on the sign and magnitude of $h(e^+)$ that is we have $f^+(\text{L})$, $f^+(\text{D})$ for $h(e^+)>0$ and $f^-(\text{L})$, $f^-(\text{D})$ for $h(e^+)<0$. As in radiolysis, $f^{\pm}(\text{D}) = f^{\mp}(\text{L})$. The asymmetry A_{Ps} in the triplet Ps formation fraction on either reversal of $h(e^+)$ ($h(e^+) \to -h(e^+)$) or interchange of L and D isomers is:

$$A_{Ps} \equiv \frac{f^+(\text{L}) - f^-(\text{L})}{f^+(\text{L}) + f^-(\text{L})} = \frac{f^+(\text{L}) - f^+(\text{D})}{f^+(\text{L}) + f^+(\text{D})} = |h(e^+)| H_{Ps}(Z) \qquad (2)$$

Symmetric expressions for A_{Ps} based on interchange of $f^+(\text{L}) \to f^-(\text{D})$ and so on may also be written. Here $H_{Ps}(Z)$ is directly analogous to $H_R(E, Z)$, but without the E dependence since Ps formation in a solid occurs in a narrow energy range below the molecular ionization potential. A detailed calculation of $H_{Ps}(Z)$ (ref. 3) shows that in our experimental conditions $H_{Ps}(Z)$ is expected to satisfy the inequality $10^{-6} < H_{Ps}(Z) < 10^{-5}$, for $Z = 6$, although in atypical cases values as large as $H_{Ps}(Z) \sim 10^{-2}$ have not been ruled out. As in the calculation of H_R, the major uncertainty in $H_{Ps}(Z)$ arises from uncertainty in the molecular wave functions.

An approximate general expression[3] relating H_R to H_{Ps} may be written as

$$H_R(E, Z) = H_{Ps}(Z) \left(\frac{\eta'_R}{\eta_{Ps}}\right) \left(\frac{1}{2E_p \ln(2E_p)}\right) \qquad (3)$$

where η_{Ps} is a molecular asymmetry factor for $H_{Ps}(Z)$, analogous to η'_R for radiolysis. The value of η_{Ps} has been calculated for the simplest optically active molecule for which the most complete wavefunctions are available, 15° twisted ethylene[3], with the result $\eta_{Ps} \sim 10^{-2}$. While η'_R has not been so calculated, it is estimated to be of the same order of magnitude as η_{Ps} (although the relative signs are not yet determined). Thus we find that $|H_R| \simeq 10^{-5} |H_{Ps}(Z)|$. Our goal is to measure A_{Ps} and $h(e^+)$ and thus determine or set upper limits on $H_{Ps}(Z)$ so that the quantities of fundamental chemical and biological interest, $h(\mathbf{\bar{r}})$ and $H_R(E, Z)$, may be estimated.

Experiment

The apparatus is designed to measure the amount of triplet Ps formed when an amino acid powder target is bombarded with a collimated beam of positrons whose energy and helicity can be varied. The asymmetry in triplet Ps production (A_{Ps}) under reversal of either positron helicity (+, −) or target chirality (L, D) is then determined. The apparatus used to produce the beam is sketched in Fig. 1 and discussed in detail in ref. 12.

Figure 1 shows the positrons focused onto the surface of a channel electron multiplier (CEM) coated with the optically active material under investigation, where their impact is detected through emission of secondary electrons (see Fig. 2). Detection of one or more of the annihilation γs in Pilot B plastic scintillator detectors (Fig. 1) provides the 'stop' signal following

the CEM 'start' signal which allows the lifetime of each positron event to be measured directly. The lifetime spectrum so obtained consists of an ~1 ns wide 'prompt peak' of directly annihilating positrons together with Ps which annihilates within the amino acid, the 140 ns exponential component of free triplet Ps, and a uniform background from uncorrelated events. To correct for variations in beam intensity, a ratio of counts from triplet events (background subtracted), occuring in a 400 ns wide time window starting at 75 ns after the prompt peak ($t = 0$), to counts in the prompt peak is calculated. Defining r^+ to be this ratio for incident positron helicity positive (r^- for $(h(e^+) < 0)$ A_{Ps} (equation (2)) is given by

$$A_{Ps} = (r^+ - r^-)/(r^+ + r^-). \tag{4}$$

Asymmetry measurements were made by reversing $h(e^+)$ every 330 s during a 12-h period. The cumulative lifetime spectrum for each value of $h(e^+)$ was stored in memory halves of a multichannel analyser. The large number of helicity reversals per run reduced the effect of random and secular time shifts in the $t = 0$ position of the prompt peak to <0.01 ns. Asymmetries were measured in this manner for samples of D, L, and DL isomers of tryptophan, cystine and leucine (supplied by Sigma Chemical Co.).

Variations in the measured asymmetries for preliminary runs were found to be greatly in excess of statistical fluctuations. Experiments showed that the Ps formation fraction could vary by as much as 20% over the surface of the CEM cone, depending on how far from the channel entrance the positron strikes the cone (see Fig. 2). As a result, changes in the size, shape, and position of the positron beam under helicity reversal could produce false asymmetries as large as 1%. Such systematic effects could be separated from any true isomeric asymmetry by noting that the time-of-flight of the secondary electrons ejected from the cone (and hence the observed time difference, $\Delta\tau$, in the prompt peak after helicity reversal), has also been shown to depend on how far from the channel entrance the beam strikes the CEM. Thus the observed asymmetries A_{Ps} with their correlated values of $\Delta\tau$ could be fit to the equation $A_{Ps} = m\Delta\tau + A_{Ps}(0)$ where m and $A_{Ps}(0)$ are the fitted parameters; and $A_{Ps}(0)$, the residual asymmetry at $\Delta\tau = 0$, is the quantity of interest. An example of the fit is shown in Fig. 3 and the results for the isomers of cystine, tryptophan and leucine are shown in Table 1.

The uncertainties in the results are statistical uncertainties for the fitted data. As evidence by the χ^2 test the linear hypothesis is statistically compatible with the data for the D, L and DL samples of cystine and tryptophan. However, using the original statistically assigned uncertainty, the data of D and L leucine yielded χ^2/N in excess of 2. In view of the still incompletely understood nature of the systematic effects discussed above, we increased the uncertainty on the individual data points to make $\chi^2/N \approx 1$. The uncertainty in $A_{Ps}(0)$ and χ^2/N (leucine) in Table 1 reflect this adjustment.

The measured values of $A_{Ps}(0)$ (Table 1) show no statistically significant difference between the D and L isomers of cystine and tryptophan at the 7×10^{-4} level but the data do indicate the possibility of an effect at the 2σ level in leucine. In view of this possible effect additional data on leucine were obtained in which an attempt was made to maintain $\Delta\tau$ at a sufficiently low level ($\Delta\tau < (0.02 \pm 0.02)$ns) so that the values of A_{Ps} from several independent runs could simply be averaged without relying on the fitting procedure described above. This was accomplished by adjusting the beam deflection electrodes before each run. The weighted averages and values of χ^2 per degree of freedom obtained using this procedure were $A_{Ps}(D) = -(30.0 \pm 4.9) \times 10^{-4}$, $A_{Ps}(L) = +(1.3 \pm 4.4) \times 10^{-4}$, $\chi^2(D)/N = 5.6/6$ and $\chi^2(L)/N = 11/7$.

The above result seems to indicate the detection of an L/D asymmetry in triplet Ps formation in leucine of $(31 \pm 7) \times 10^{-4}$. However, in view of the large systematic effect discussed previously (values of A_{Ps} of up to -60×10^{-4} were observed in leucine for $\Delta\tau \sim 0.1$ ns) and in view of the fact that $A_{Ps} \sim 30 \times 10^{-4}$ is a much larger effect than predicted by theory[3], we feel that this result is still preliminary, and while suggestive of an effect, it is not definitive. A newly designed experiment is under construction in which the systematic effects related to beam parameters and detector geometry should be eliminated. We conclude that we have established A_{Ps} to be $<7 \times 10^{-4}$ in cystine and tryptophan, but that there may be an effect in leucine. We note that previous work by Bonner *et al.*[7] did show asymmetric L/D radiolysis by polarized electrons in leucine but that this effect was not corroborated by Hodge *et al.*[8].

To relate A_{ps} to $H_{Ps}(Z)$ (equation (2)) it is necessary to obtain the positron helicity $(h(e^+) = \langle\hat{\mathbf{s}}_i\rangle \cdot \langle\hat{\mathbf{p}}_i\rangle)$ not for the incident positron beam (recall $h_0(e^+) = 0.21$), but for the beam after it has slowed down in the target to Ps formation energies (≤ 10 eV). This final helicity ($h_f(e^+)$) is not directly observable but it can be estimated theoretically. Many experiments[12,13] have shown that, in agreement with theoretical predictions[14], the magnitude and direction of $\langle\hat{\mathbf{s}}\rangle$ for an ensemble of positrons is essentially unchanged when the positrons slow down to Ps formation energies from energies far in excess of the initial 400 eV energy used in this experiment. Although $\langle\hat{\mathbf{s}}_i\rangle$ is constant, the magnitude of $\langle\hat{\mathbf{p}}_i\rangle$ $(|\langle\hat{\mathbf{p}}_i\rangle|)$ does decrease as the beam slows down and this increased beam divergence can be calculated for positrons of the energies we use incident on metal surfaces[15,16]. For the case of $|\langle\hat{\mathbf{p}}_i\rangle|_{\text{initial}} = 1$, the beam divergence after n collisions is given by $|\langle\hat{\mathbf{p}}_i\rangle| = \cos\bar{\theta} = [\cos\theta_1 \cos\theta_2 \cdots \cos\theta_n]$. The theoretical results are generally in good qualitative agreement with experiment[17,18].

The calculation of $\cos\bar{\theta}$ using the formulas of refs 17, 18 may be generalized to the case of the amorphous insulator alumina (Al_2O_3) based on calculations of specific energy loss (dE/dx) and mean free path between collisions (λ)[19]. The final extrapolation to the amino acids used in our experiment is made based on the assumption that dE/dx and λ may be scaled from one substance to another because the probability of a positron–atom interaction is proportional to $N\sigma$ where N is the number of electrons/atom with which the positron can interact (essentially valence electrons) and σ is the positron–atomic electron interaction cross-section (taken as constant in all amorphous insulators).

The results of the calculation yield a final value of $\cos\bar{\theta} = (0.4 \pm 0.2)$ or $h_f(e^+) = h_0(e^+)|\langle\hat{\mathbf{p}}_i\rangle| \approx 0.2 \times 0.4 = (0.08 \pm 0.04)$. The error assignment results from systematic uncertainties in the correct number of valence electrons to use in a given amino acid and the nature of various approximations of the scattering cross-sections used in ref. 19.

Conclusion

Our experiment aims to determine $H_{Ps}(Z)$, or set limits on it, so that the quantity of interest relating to the origin of optical activity $H_R(E, Z)$ may be determined (equation (3)), and so that the newly identified property of chiral molecules, the helicity density $h(\mathbf{\bar{r}})$, may be investigated. From equation (2) and the value $h_f(e^+) \simeq 0.08 \pm 0.04$ we have $H_{Ps}(Z) \leq (A_{Ps}/h(e^+)) = (7 \times 10^{-4}/8 \times 10^{-2}) \simeq 10^{-2}$ for cystine and tryptophan. The possible value of $H_{Ps}(Z)$ for leucine where an effect of $A_{Ps} = (31 \pm 7) \times 10^{-4}$ may have been observed is $H_{Ps}(Z) \simeq (31 \times 10^{-4}/8 \times 10^{-2}) \sim (4 \pm 2) \times 10^{-2}$.

Comparing the above results with previous measurements[20–23] using positrons we note that, in all cases, positrons from a radioactive source slowed to positronium formation energies directly in the amino acid, implying[9] $h_f(e^+) < 10^{-2}$. Three experiments[20–22] gave null values of A_{Ps} at the 10^{-2} level, which if combined with the estimated $h_f(e^+)$, yield $H_{ps}(Z) < 10$. The positive results of ref. 23 imply $H_{Ps}(Z) \geq 100$. Thus the results of the present experiment, implying $H_{ps}(Z) \leq 10^{-2}$ in three cases and $H_{Ps}(Z) = 4 \times 10^{-2}$ in the fourth, although larger than the theoretical predictions of ref. 3 ($H_{Ps}(Z) < 10^{-5}$), are the first to set limits on $H_{Ps}(Z)$ which have any reasonable physical meaning, that is $H_{Ps}(Z) < 1$. Finally, using equation (3), our limits on $H_{Ps}(Z)$ enable us to set the upper limit

$H_R(E,Z) < 10^{-7}$. This represents an improvement of 10^4 over the limits of 10^{-3} set by direct radiolysis experiments.

Plans for future experiments involve changes to reduce the instrumental asymmetries and to increase the beam rate by an order of magnitude. In addition, we are now able to increase beam polarization up to 0.7 (J.V.H and P.W.Z in preparation). These and other changes should enable us to improve our limits on H_{Ps} by a further factor of ~ 100, sufficient to observe an effect if $H_{Ps}(Z) \sim 10^{-4}$, a value within theoretical bounds. Finally, the predicted[3] increase in asymmetry with Z^2 will be tested using amino acids with heavy atoms ($Z > 30$) at the chromophore. Observation of the asymmetry in such a molecule should be within reach of the next generation experiment.

We thank Professors G. W. Ford, R. A. Hegstrom, G. Karl, R. R. Lewis, and J. Walker for helpful discussions and R. H. Sands for major equipment loans. This work is supported by NASA and by the NSF.

Received 30 December 1981; accepted 7 April 1982.

1. Kizel, V. A. Soviet Phys. Usp. 23, 277–295 (1980).
2. Norden, B., J. molec. Evolut. 11, 313–332 (1978).
3. Hegstrom, R. A. Nature 297, 643–647 (1982).
4. Norden, B. Nature 266, 567–568 (1977).
5. Angel, J. R. P. & Illing, R. Nature 238, 389–390 (1972).
6. Keszthelyi, L. Origins of Life 8, 299–340 (1977).
7. Bonner, W. A., Van Dort, M. A. & Yearian, M. A. Nature 258, 419–421 (1975).
8. Hodge, L. A., Dunning, F. B. & Walters, G. W. Nature 280, 250–252 (1979).
9. Rich, A. Nature 264, 482 (1976).
10. Farago, P. J. Phys. B 13, L567–569 (1980); 14, L743–748 (1981).
11. Beerlage, M. J. M., Farago, P. S. & Van der Wiel, M. J. J. Phys. B 14, 3245–3253 (1981).
12. Gerber, G., Newman, D., Rich, A. & Sweetman, E. Phys. Rev. D 15, 1189–1193 (1977).
13. Zitzewitz, P. W., Van House, J. C., Rich A. & Gidley, D. W. Phys. Rev. Lett. 43, 1281–1284 (1979).
14. Bouchiat, C. & Levy-Leblond, L. M. Nuovo Cim. 33, 193–200 (1964).
15. Oliva, J. Phys. Rev. B 21, 4909–4924 (1980).
16. Nieminen, R. M. & Oliva, J. Phys. Rev. B 22, 2226–2247 (1980).
17. Lanteri, H., Bindiet, R. & Rostaing, P. Thin Solid Films 67, 293–299 (1980).
18. Hrach, R. Thin Solid Films 15, 65–69 (1973).
19. Ritchie, R. H., Garber, F. W., Nakai, M. Y. & Birkhoff, R. D. Adv. Radiat. Biol. 3 (1969).
20. Dezsi, I., Horvath, D. & Kajcsos, Z. S. Chem. Phys. Lett. 24, 514–515 (1974).
21. Brandt, W. & Chiba, T. Phys. Lett. 57A, 395–396 (1977).
22. Jean, Y. C. & Ache, H. J. J. chem. Phys. 81, 1157–1162 (1977).
23. Garay, A. S., Keszthelyi, L., Demeter, I. & Hrasko, P. Nature 250, 332–333 (1974); Chem. Phys. Lett. 23, 549–552 (1973).
24. Gidley, D. W., Rich, A., Van House, J. C. & Zitzewittz, P. W. Origins of Life 11, 31–36 (1981); in Origin of Life, (ed. Wohlman, Y.) 379–384 (Reidel, Boston, 1981).
25. Rich, A., Van House, J. & Hegstrom, R. A. Phys. Rev. Lett. 48, 1341–1344 (1982).

β Decay and the origins of biological chirality: theoretical results

Roger A. Hegstrom

Department of Chemistry, Wake Forest University, Winston-Salem, North Carolina 27109, USA

A dynamical mechanism is found whereby a dissymmetric molecule and its mirror image are ionized at different rates by longitudinally polarized electrons such as produced by nuclear β decay. An enhancement is predicted for molecules containing heavy atoms. Order-of-magnitude estimates indicate that the asymmetric effect of this mechanism may be detectable by current experiments on positronium formation.

THE postulated existence of a causal relationship between the parity nonconserving aspect of the weak interaction and the observed dissymmetry of the molecules present in living organisms[1-3] seems reasonable as the weak interaction is universal and always acts in the same chiral sense. Hence if an effective mechanism exists involving the weak interaction in molecular evolution, this inherent universal chirality in nature could have influenced the selection of exclusively D sugars for RNA and DNA and L amino acids for proteins.

Two candidates for such a mechanism have been proposed, each of which involves a different aspect of the weak interaction. One produces an energy difference between a chiral molecule and its mirror image[4], but the calculated difference[5] seems too small to have been effective in molecular evolution. The other, which seems the most popular theory at present, postulates the asymmetric radiolysis of racemic mixtures of prebiotic chiral molecules by the longitudinally polarized electrons produced in nuclear β decay[6]. A precise test of some aspects of this mechanism is the object of new experiments to measure an asymmetry in the rate of triplet positronium (Ps) formation in chiral molecules discussed in the accompanying article[7] (where references to previous work may be found).

The present article gives a theoretical basis for, and estimates the magnitudes of, the cross-section asymmetries for both Ps formation and radiolysis. The relevance of the results to the question of the origin of chirality in living organisms is discussed extensively in the preceding article[7].

Ps formation in optically active molecules

The treatment of the asymmetry in the cross-section for positronium formation given here and that for radiolysis which follows are based on nonrelativistic scattering theory. The notation of Taylor[7] is followed except that atomic units are used here ($h/2\pi = e = m_e = 1$, where h is Planck's constant, e the electron charge, and m_e the electron mass).

Consider a positron with momentum $\vec{\mathbf{p}}$ and helicity $\lambda = \vec{\sigma} \cdot \hat{\mathbf{p}} = \pm 1$, where $\hat{\mathbf{p}} = \vec{\mathbf{p}}/p$ and $p = |\vec{\mathbf{p}}|$, incident on an optically active molecule, the nuclei of which are taken to be fixed in the laboratory frame of reference (Fig. 1). This initial channel is labelled α, and α' labels the final channel which consists of the ionized molecule and a Ps atom with momentum $\vec{\mathbf{P}}'$. Spin is quantized along $\vec{\mathbf{p}}$ and hence the spin projection quantum number of the positron is related simply to the helicity by $m_\lambda = \frac{1}{2}\lambda$. The cross-section depends on the helicity λ, the chirality of the target molecule (L or D), and on the spin state S of the Ps, and is given by

$$\sigma_S^\lambda(\mathrm{L}) = (2\pi)^4 \frac{P'}{p} \int |t_S^\lambda(\mathrm{L})|^2_{\mathrm{Av}}\, d\Omega' \tag{1}$$

(with a corresponding expression for the D isomer) where $P' = |\vec{\mathbf{P}}'|$, where $d\Omega'$ denotes an integration over the solid angles corresponding to directions of $\vec{\mathbf{P}}'$, and where Av denotes an average over all orientations of the molecule (random orientations are assumed). The quantity t_S^λ, which is related to the on-shell T-matrix, is defined below.

In a notation consistent with that of the preceding paper[7], the asymmetry in the cross-section is defined as

$$H_{Ps}(\mathrm{L}) \equiv \frac{\sigma_S^+(\mathrm{L}) - \sigma_S^-(\mathrm{L})}{\sigma_S^+(\mathrm{L}) + \sigma_S^-(\mathrm{L})} = \frac{\sigma_S^+(\mathrm{L}) - \sigma_S^+(\mathrm{D})}{\sigma_S^+(\mathrm{L}) + \sigma_S^+(\mathrm{D})} \tag{2}$$

(with a corresponding equation for the D isomer) where $\sigma_S^+(\mathrm{L})$ and $\sigma_S^-(\mathrm{L})$ are obtained from equation (1) with $\lambda = +1$ and $\lambda = -1$ respectively. The second equation in equation (2) is a consequence of parity conservation for the electromagnetic interaction [equations (6, 7)], from which the general relationship $\sigma_S^\lambda(\mathrm{L}) = \sigma_S^{-\lambda}(\mathrm{D})$ follows. For the case of a positron beam of helicity $h(e^+)$ and with $h \equiv |h(e^+)| \leq 1$, it can be shown that, for a given L or D isomer (from now on the isomer is not denoted explicitly), the asymmetry in the cross-section obtained

by reversing the beam helicity is

$$A_{Ps} \equiv \frac{\sigma_S^{+h} - \sigma_S^{-h}}{\sigma_S^{+h} + \sigma_S^{-h}} = hH_{Ps} \tag{3}$$

The dynamics which produce the cross-section asymmetry are now considered. For simplicity, a single-electron treatment of the molecular bound state is given here and illustrated in Fig. 1. The positron, 'a', is incident on a molecule consisting of an electron 'b' bound to a core 'c'. The core approximates the fixed nuclear framework and the other electrons in the molecule and provides an effective potential for b. The spin of b, like that of a, is quantized along $\hat{\mathbf{p}}$ with its direction given by $s \equiv 2m_s = \pm 1$, where m_s is the spin projection quantum number. The electron bound state is denoted ψ_s.

The cross-section is given by equation (1), where, for a spin-unpolarized target molecule

$$|t_S^\lambda|^2 = \tfrac{1}{2}\sum_{m_s}\sum_{M_S}|t_{SM_S}^{m_\lambda m_s}|^2 \tag{4}$$

where the second summation is taken over the spin projection quantum number M_S corresponding to the spin S. The on-shell T-matrix is given by

$$t_{SM_S}^{m_\lambda m_s} = \langle \vec{\mathbf{P}}'\phi_{Ps} SM_S - |V^\alpha|\vec{\mathbf{p}}m_\lambda\psi_s\rangle \tag{5}$$

where ϕ_{Ps} denotes the spatial part of the Ps wave function, V^α denotes the potential energy giving the interaction between the incident positron and the molecule, and the minus sign denotes the prior form[8] of the T-matrix.

It can be shown on the basis of equations (1)–(5) that H_{Ps} vanishes unless (1) the molecule is dissymmetric and (2) spin-dependent interactions are involved. According to equation (5) there are three possible ways for the effects of spin-dependent forces to occur: (1) in the interaction potential V^α, (2) in the positronium wave function, and (3) in the molecular wave function ψ_s. An investigation of each of these possibilities indicates that for molecules containing atoms as large as carbon, or larger, the dominant contribution to the asymmetry H_{Ps} is expected to come from the perturbation of the electron bound state ψ_s by the spin–orbit interaction of the bound electron moving in the electric field of the molecular core. In what follows, only this bound-state spin–orbit interaction is kept and all other spin-dependent terms in the hamiltonian are ignored. The resulting description is called the bound helical electron model, because, as will be seen later the theory predicts the existence of a helicity density for the electrons bound in an optically active molecule. The qualitative aspects of this model resemble the "helical electron gas" model of Hrasko and Garay[9].

The dynamics of the bound helical electron model are expressed in the total Hamiltonian $H = H^\alpha + V^\alpha$ for the system, where the channel α Hamiltonian is given by

$$H^\alpha = T_a + T_b + V_{bc}^{COUL} + V_{bc}^{SO} \tag{6}$$

and the interaction V^α by

$$V^\alpha = V_{ab}^{COUL} + V_{ac}^{COUL} \tag{7}$$

where the subscripts on the kinetic energy operators T and potential energy operators V denote the particles a, b and core c, and where 'COUL' and 'SO' denote the Coulomb and spin–orbit interactions, respectively. Now the asymptotic stationary state $|\vec{\mathbf{p}}m_\lambda\psi_s\rangle$ appearing in equation (5) is an eigenfunction of H^α with eigenvalue E, the total energy of the system, so that when the spin–orbit energy is small compared with E first-order perturbation theory may be used to obtain the result

$$|\psi_s\rangle = \left(\phi_0 + s\sum_n \varepsilon_n^0\phi_n\right)|m_s\rangle + \sum_n \varepsilon_n^s\phi_n|\bar{m}_s\rangle \tag{8}$$

where $\phi_0|m_s\rangle$ is the unperturbed molecular bound state, $\phi_n|m_s\rangle$ are excited states, $\bar{m}_s = -m_s$, and

$$V_{bc}^{SO} = \vec{\mathbf{\Lambda}}_{bc}\cdot\vec{\boldsymbol{\sigma}}_b \equiv \tfrac{1}{4}\alpha^2\vec{\mathbf{E}}_{bc}\times\vec{\mathbf{p}}_b\cdot\vec{\boldsymbol{\sigma}}_b \tag{9a}$$

$$\vec{\boldsymbol{\varepsilon}}_n = (E_0 - E_n)^{-1}\langle\phi_n|\vec{\mathbf{\Lambda}}_{bc}|\phi_0\rangle \tag{9b}$$

$$\varepsilon_n^0 = \vec{\boldsymbol{\varepsilon}}_n\cdot\hat{\mathbf{e}}_3 = \vec{\boldsymbol{\varepsilon}}_n\cdot\hat{\mathbf{p}} \tag{9c}$$

$$\varepsilon_n^\pm = \vec{\boldsymbol{\varepsilon}}_n\cdot\hat{\mathbf{e}}_\pm = \vec{\boldsymbol{\varepsilon}}_n\cdot(\hat{\mathbf{e}}_1 \pm i\hat{\mathbf{e}}_2) \tag{9d}$$

Here $\alpha \sim 137^{-1}$ is the fine structure constant, $\vec{\mathbf{E}}_{bc}$ is the electric field produced by the core c at the location of electron b, $\vec{\mathbf{p}}_b$ is the momentum of b, $\vec{\boldsymbol{\sigma}}_b$ is the Pauli spin matrix, and $\hat{\mathbf{e}}_k$ are unit vectors with $\hat{\mathbf{e}}_3 = \hat{\mathbf{p}}$.

The asymmetry of the positronium formation cross-section is now considered within the bound helical electron model. It can be shown that the 'spin-flip' term in the perturbed molecular wave function, that is, the last term of equation (8), does not contribute to the cross-section to first-order in the spin–orbit interaction and hence can be ignored. The contributing part of equation (8) is now abbreviated $\phi_s|m_s\rangle$, where

$$\phi_s = \phi_0 + s\sum_n \varepsilon_n^0\phi_n \tag{10}$$

Substitution of $\phi_s|m_s\rangle$ for $|\psi_s\rangle$ in equation (5) then gives

$$t_{SM_S}^{m_\lambda m_s} = V^s\langle SM_S|m_\lambda m_s\rangle \tag{11}$$

where

$$V^s = \langle\vec{\mathbf{P}}'\phi_{Ps} - |V^\alpha|\vec{\mathbf{p}}\phi_s\rangle \tag{12}$$

The spin-dependent factors appearing in equation (11) are easily evaluated to obtain the contributing T-matrix elements in terms of V^s, and then equations (1)–(4) are used to obtain the cross-section asymmetry H_{Ps} or A_{Ps} to first order in the spin–orbit interaction. For the triplet state the result is

$$H_{Ps} = \frac{1}{6}\frac{\int\{|V^+|^2_{Av} - |V^-|^2_{Av}\}\,d\Omega'}{\int|V^0|^2_{Av}\,d\Omega'} \tag{13}$$

where $V^\pm$ are defined by equation (12) and where V^0 is obtained by replacing ϕ_s by ϕ_0 on the right-hand side of equation (12). The result for the singlet state is obtained by replacing the factor 1/6 in equation (13) by −1/2.

The precise calculation of H_{Ps} is an extremely difficult collision problem. Its order of magnitude can be estimated simply, however, using equations (9)–(13), from which it follows that H_{Ps} is first order in the spin–orbit coefficient ε_n. On the basis of dimensional arguments H_{Ps} would then be expected to be of order $|\varepsilon_n| \sim (\alpha Z)^2 = (Z/137)^2$ where Z is the electric charge on the heaviest atom or atoms in any asymmetric environment in the molecule[4]. Following the treatment of the parity nonconserving energy difference[4], H_{Ps} is expressed

$$H_{Ps} = \eta_{Ps}(\alpha Z)^2 \tag{14}$$

where the molecular asymmetry factor η_{Ps} is included to take into account the effect of the molecular environment on the magnitude of H_{Ps}. An approximate calculation of H_{Ps} (R.H. and R. Stewart, unpublished data; see also ref. 5) for the simplest dissymmetric molecules for which a wave function has been published, namely twisted ethylene, suggests $\eta_{Ps} \sim 10^{-2}$ to 10^{-3}, which may be typical, although values outside this range cannot be ruled out until more extensive calculations have been performed.

At present the Michigan experiments[7] have examined three amino acids, two of which give the result $|H_{Ps}| < 10^{-2}$ in agreement with the present theoretical predictions (taking $\eta_{Ps} = 10^{-2}$

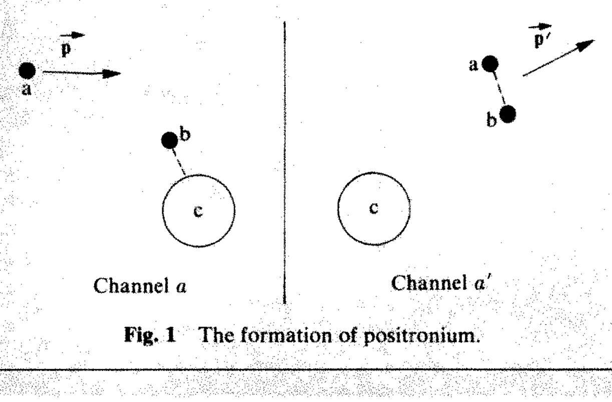


Fig. 1 The formation of positronium.

to 10^{-3} gives $H_{Ps} \sim 10^{-5}$–10^{-6} for amino acids). The third amino acid, leucine, gives the preliminary result $H_{Ps} = (4 \pm 2) \times 10^{-2}$ in apparent disagreement with theory. Although it is conceivable that the present theory could obtain such a large value of H_{Ps} in atypical cases, it is premature to attempt to make a more detailed analysis for the case of leucine at present due to the difficulty of even crude calculations involving amino acid wave functions and also due to the preliminary nature of the experimental result.

Finally, in view of the treatment of radiolysis to be given next, it should be noted that the asymmetry in the total inclusive cross-section for Ps formation, obtained by summing equation (1) over $S = 0$ and 1, vanishes identically due to the cancellation of the singlet and triplet contributions. A vanishing result is also obtained for ionization without Ps formation, which is closely related to the process of ionization by polarized electrons treated next.

Electron impact ionization of optically active molecules (radiolysis)

The theoretical treatment of radiolysis is similar to that just given for Ps formation. The most important difference is that, because the electrons are identical, the two-electron state of incident plus target electron is antisymmetric with respect to the interchange of these particles, and as a consequence both direct and exchange terms appear. In fact it is the existence of these exchange terms which leads to a nonvanishing asymmetry in the total cross-section, as is shown below, in contrast to the case of ionization by positrons discussed above.

The scattering process is illustrated in Fig. 2. In the initial channel α the incident electron 1 has helicity $\lambda = \pm 1$ and momentum $\vec{\mathbf{p}}$ and the target electron 2 is bound to the core c. In the two final channels α' and $\tilde{\alpha}'$, which differ by an exchange of the two electrons, the momenta are designated $\vec{\mathbf{p}}_a$ and $\vec{\mathbf{p}}_b$. The spins are again quantized along $\hat{\mathbf{p}}$ with initial spin projection quantum numbers m_λ, m_s and final spin quantum numbers S, M_S describing singlet and triplet states. The total ionization cross-section, inclusive of electrons in singlet and triplet final states, is given (for either the D or L isomer) by

$$\sigma^\lambda = \tfrac{1}{2}(2\pi)^4 \frac{1}{p} \int_0^{E_p - I} \mathrm{d}E_b p_a p_b \int \mathrm{d}\Omega_a \, \mathrm{d}\Omega_b |\hat{\mathbf{t}}^\lambda|^2_{Av} \quad (15)$$

where $E_p = p^2/2$ is the energy of the incident electron, I is the molecular ionization energy, $E_b = p_b^2/2$, and where

$$|\hat{\mathbf{t}}^\lambda|^2 = \tfrac{1}{2} \sum_{m_s} \sum_S \sum_{M_S} |\hat{\mathbf{t}}^{m_\lambda m_s}_{SM_S}|^2. \quad (16)$$

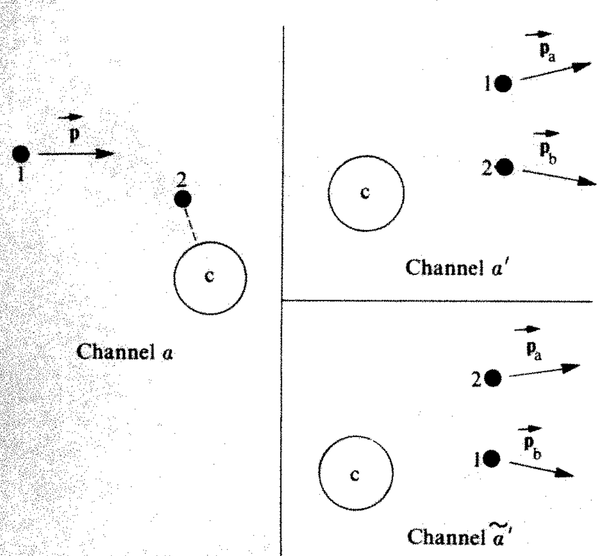


Fig. 2 Ionization by electron impact.

The on-shell T-matrix elements are given by

$$\hat{\mathbf{t}}^{m_\lambda m_s}_{SM_S} = t^{m_\lambda m_s}_{SM_S}(\mathrm{dir}) + (-)^S t^{m_\lambda m_s}_{SM_S}(\mathrm{exch}) \quad (17)$$

with the direct and exchange terms given by

$$t^{m_\lambda m_s}_{SM_S}(\mathrm{dir}) = \langle \vec{\mathbf{p}}_a \vec{\mathbf{p}}_b SM_S - |V^\alpha| \vec{\mathbf{p}} m_\lambda \psi_s \rangle \quad (18a)$$

$$t^{m_\lambda m_s}_{SM_S}(\mathrm{exch}) = \langle \vec{\mathbf{p}}_b \vec{\mathbf{p}}_a SM_S - |V^\alpha| \vec{\mathbf{p}} m_\lambda \psi_s \rangle \quad (18b)$$

Equations (18a, b) are analogous to equation (5) and ψ_s is again given by equation (8). The same procedure which led from equation (5) to equation (13) is again followed here using equations (15)–(18) and gives the following result for the asymmetry in the radiolysis cross-section, correct to first order in the spin–orbit interaction:

$$H_R = \frac{\sigma^+ - \sigma^-}{\sigma^+ + \sigma^-} = -\frac{1}{2} \frac{\int_0^{E_p - I} \mathrm{d}E_b p_a p_b \int \mathrm{d}\Omega_a \, \mathrm{d}\Omega_b \, Re\{V_d^{+*} V_e^+ - V_d^{-*} V_e^-\}_{Av}}{\int_0^{E_p - I} \mathrm{d}E_b p_a p_b \int \mathrm{d}\Omega_a \, \mathrm{d}\Omega_b \{|V_d^0|^2 + |V_e^0|^2 - Re V_d^{0*} V_e^0\}_{Av}} \quad (19)$$

where

$$V_d^s = \langle \vec{\mathbf{p}}_a \vec{\mathbf{p}}_b - |V^\alpha| \vec{\mathbf{p}} \phi_s \rangle \quad (20a)$$

$$V_e^s = \langle \vec{\mathbf{p}}_b \vec{\mathbf{p}}_a - |V^\alpha| \vec{\mathbf{p}} \phi_s \rangle \quad (20b)$$

are direct and exchange integrals with ϕ_s again given by equation (10). The integrals V_d^0 and V_e^0 are obtained by replacing ϕ_s with ϕ_0 in equations (20a, b). If the incident electron state is a mixture of pure helicity states with a net helicity $h(e^-)$, then the asymmetry in the cross-section is

$$A_R = \frac{\sigma^{+h} - \sigma^{-h}}{\sigma^{+h} + \sigma^{-h}} = hH_R \quad (21)$$

where $h = |h(e^-)|$. A negative value of H_R or A_R for a given L or D isomer implies that the isomer is preferentially ionized by natural β electrons, for which $h(e^-)$ is negative.

Although the precise calculation of H_R is difficult, it is possible to make rough estimates of its order of magnitude as was done for H_{Ps}. It is useful to distinguish two cases. For the case of incident energies E_p near the ionization threshold, the final momenta $\vec{\mathbf{p}}_a$, $\vec{\mathbf{p}}_b$ both approach zero, so it follows from equations (20a, b) that $V_e^s \to V_d^s$ (note that only scattering into the singlet state contributes in this limit), and then from equation (19) that

$$H_R \xrightarrow[E_p \to I]{} -\frac{1}{2} \frac{|V_d^+|^2_{Av} - |V_d^-|^2_{Av}}{|V_d^0|^2_{Av}} \quad (22)$$

which, on the basis of dimensional arguments similar to those which led to equation (14), can be shown to have the same order of magnitude as H_{Ps}. Hence H_R is expressed

$$H_R = \eta_R (\alpha Z)^2 \qquad (E_p \to I) \quad (23)$$

where the molecular asymmetry factor for radiolysis η_R is expected to have the same order of magnitude as η_{Ps}. For the case of incident energies near 100 keV, which is typical of β-decay electrons, use of the nonrelativistic Bethe approximation[10] in equation (19) gives the result

$$H_R = \frac{\eta'_R (\alpha Z)^2}{(2E_p) \ln (2E_p)} \qquad (1 \ll E_p \ll \alpha^{-2}) \quad (24)$$

where η'_R is formally of the order of one but is expected to be less than one for the same reasons given in connection with equations (14) and (23). The range for E_p corresponds to the range of validity for the nonrelativistic Bethe approximation[10]; for more precise calculations relativistic effects should be included. For $E_p = 100$ keV $\simeq 3{,}700$ AU equation (24) gives $H_R \simeq 10^{-5} \eta'_R (\alpha Z)^2$ which is roughly five orders of magnitude smaller than the threshold value. This large reduction is due primarily to the decreasing magnitudes of the exchange integrals V_e^s with increasing energy. Taking η'_R to be of the order 10^{-2}–10^{-3} in equation (24) gives an estimated asymmetry $H_R \sim$

10^{-10}–10^{-11} for the radiolysis of amino acids ($Z=6$) with 100-keV electrons and a magnitude roughly 100 times larger for a molecule containing a dominant heavy atom with Z near 100.

Recently Zel'dovich and Saakyan[11] have calculated a contribution to H_R which is not included in equation (19) but rather is due to the direct spin-dependent interaction between the incident and target electrons, and which may be expressed

$$H_R(\text{ref. } 10) \simeq \alpha^2 \frac{v}{c} \frac{\sum_f \mathrm{Im}(\vec{\mathbf{d}}_{fi} \cdot \vec{\mathbf{m}}_{fi}^*)}{\sum_f |\vec{\mathbf{d}}_{fi}|^2} \tag{25}$$

where v/c is the ratio of the speed of the incident electron to the speed of light, and where $\vec{\mathbf{d}}_{fi}$ and $\vec{\mathbf{m}}_{fi}$ are the electric and magnetic transition moments, respectively, between the initial and final molecular states. The ratio of sums appearing in equation (25) may be estimated from spectral data on optically active molecules[12]. For strong absorption bands, which are expected to dominate the cross-section, this ratio is typically of the order of 10^{-6}, which is much less than the estimated value 10^{-2}–10^{-3} based on dimensional arguments[11]. Equation (25) then gives the result H_R (ref. 11) $\simeq 10^{-11}$ for 100 keV electrons, which is comparable with the result for the bound helical electron model at the same energy and with $Z \simeq 6$. Hence the contribution to H_R from equation (19) dominates at low energies and high Z, but the contributions from both equations (19) and (25) are important at high energies.

More recently, Mann and Primakoff[13] have estimated H_R to be relatively large (of the order of $\alpha^2 \approx 10^{-4}$) due to the interaction between the spin of the incident beta electron and the orbital angular momentum of the target electron. This estimate is in direct disagreement with that from ref. 12. My preliminary calculation gives the result that the contribution to H_R from this spin–other-orbit interaction, although formally of the order of α^2, vanishes in the first Born approximation when the average over molecular orientations is taken.

I conclude this section with the following remarks: (1) β Electrons are emitted with a spectrum of energies ranging from zero up to a maximum which is typically hundreds of keV. In view of the present results, the low energy end of this spectrum may be more important for asymmetric radiolysis than previously thought, even though the magnitude of the helicity is less at low energies. (2) Experiments which attempt to detect this asymmetry are expected to find a larger effect for low-energy polarized electrons, particulary with dissymmetric target molecules containing heavy atoms.

Electron helicity in bound states

For a one-electron state ψ_s which is characterized by the spin direction $s \equiv 2m_s$ the helicity density is defined in coordinate space by the equation

$$h_s(\vec{\mathbf{x}}) = Re\{\psi_s^\dagger(\vec{\mathbf{x}})\vec{\boldsymbol{\sigma}} \cdot \hat{\mathbf{p}}_e \psi_s(\vec{\mathbf{x}})\} \tag{26a}$$

or in momentum space by

$$h_s(\vec{\mathbf{p}}_e) = Re\{\psi_s^\dagger(\vec{\mathbf{p}}_e)\vec{\boldsymbol{\sigma}} \cdot \hat{\mathbf{p}}_e \psi_s(\vec{\mathbf{p}}_e)\} \tag{26b}$$

where $\vec{\boldsymbol{\sigma}}$ is the Pauli spin operator, $\vec{\mathbf{p}}_e$ is the electron momentum, and $\hat{\mathbf{p}}_e = \vec{\mathbf{p}}_e/|\vec{\mathbf{p}}_e|$. The two-component spin functions $\psi_s(\vec{\mathbf{x}})$ and $\psi_s(\vec{\mathbf{p}}_e)$ are Fourier transforms of each other. For a one-electron atom or molecule in a pure spin state, the helicity density can be relatively large, of the order of several atomic units.

The helicity density may also be defined in a general way for a many-electron system, but here it is sufficiently general to consider the simple case of a two-electron singlet state

$$\Psi(1,2) = 2^{-1/2}|\psi_+(1)\psi_-(2) - \psi_-(1)\psi_+(2)\rangle \tag{27}$$

for which the helicity density is given by

$$h(\vec{\mathbf{p}}_e) = \sum_{s=\pm 1} Re\{\psi_s^\dagger(\vec{\mathbf{p}}_e)\vec{\boldsymbol{\sigma}} \cdot \hat{\mathbf{p}}_e \psi_s(\vec{\mathbf{p}}_e)\} \tag{28a}$$

$$= \sum_{s=\pm 1} h_s(\vec{\mathbf{p}}_e) \tag{28b}$$

In this case it can be easily shown that, in the absence of spin–orbit coupling, the helicity density vanishes due to cancellation of equal and opposite contributions from the two spin states. If spin–orbit coupling corrections to the wave function $\Psi(1,2)$ are now included by perturbing the spin orbitals $\psi_\pm$ according to equation (8), then equation (28) gives

$$h(\vec{\mathbf{p}}_e) = 4Re \sum_n \vec{\boldsymbol{\varepsilon}}_n \cdot \hat{\mathbf{p}}_e \phi_0^*(\vec{\mathbf{p}}_e)\phi_n(\vec{p}_e) \tag{29}$$

where $\vec{\varepsilon}_n$ is defined in equation (9b). The same result (divided by two) is obtained for a spin-unpolarized one-electron state. The following comments regarding equation (29) can be made: (1) h is formally of the order of $|\varepsilon_n| \sim (\alpha Z)^2$ atomic units; (2) h vanishes for an atom or for a molecule with a centre of inversion symmetry; (3) h is in general non-zero for a molecule with a fixed orientation and without a centre of inversion; (4) if an average over orientations is taken for an ensemble of randomly oriented molecules, the resulting average helicity density h_{Av} is independent of the direction of $\hat{\mathbf{p}}_e$ and is non-zero only in the case of a dissymmetric molecule; (5) the expectation value $\langle\vec{\boldsymbol{\sigma}} \cdot \hat{\mathbf{p}}_e\rangle$, which is obtained from the right-hand side of equation (29) or equation (28a) by replacing $\hat{\mathbf{p}}_e$ with $\vec{\mathbf{p}}_e$ and integrating over all momentum space, is identically zero to first order in the spin–orbit interaction. Proofs of these statements will be given elsewhere. The helicity density is an observable property of atoms and molecules which apparently has not been considered previously.

The asymmetry in the cross-sections for triplet Ps formation in dissymmetric molecules may be calculated from the helicity density by means of an approximate equation derived from equation (13):

$$H_{Ps} \simeq \frac{1}{6} \frac{\int d\Omega' |F_{Ps}(\vec{\mathbf{Q}})|^2 \hat{\mathbf{p}} \cdot \hat{\mathbf{q}} h_{Av}(q)}{\int d\Omega' |F_{Ps}(\vec{\mathbf{Q}})|^2 |\phi_0(\vec{\mathbf{q}})|^2_{Av}} \tag{30}$$

where $F_{Ps}(\vec{\mathbf{Q}})$ is the Fourier transform of the product of the Ps wave function and an effective potential energy for the interaction between the incident and target particles, $\vec{\mathbf{q}} = \vec{\mathbf{P}}' - \vec{\mathbf{p}}$, $\vec{\mathbf{Q}} = \frac{1}{2}(\vec{\mathbf{q}} - \vec{\mathbf{p}})$, and where the other quantities have been defined previously. A similar approximate relation is obtained for the radiolysis asymmetry:

$$H_R \simeq -\frac{1}{2} \frac{\int_0^{(E_p-I)/2} dE_b p_a p_b \int d\Omega_a\, d\Omega_b F_R(\vec{\mathbf{Q}}_a) F_R(\vec{\mathbf{Q}}_b) \hat{\mathbf{p}} \cdot \hat{\mathbf{q}} h_{Av}(q)}{\int_0^{(E_p-I)/2} dE_b p_a p_b \int d\Omega_a\, d\Omega_b [F_R^2(\vec{\mathbf{Q}}_a) + F_R^2(\vec{\mathbf{Q}}_b) - F_R(\vec{\mathbf{Q}}_a) F_R(\vec{\mathbf{Q}}_b)] |\phi_0(\vec{\mathbf{q}})|^2_{Av}} \tag{31}$$

where F_R is defined similarly to F_{Ps} and where $\vec{\mathbf{Q}}_a = \vec{\mathbf{p}}_a - \vec{\mathbf{p}}$, $\vec{\mathbf{Q}}_b = \vec{\mathbf{p}}_b - \vec{\mathbf{p}}$, and $\vec{\mathbf{q}} = \vec{\mathbf{p}}_a + \vec{\mathbf{p}}_b - \vec{\mathbf{p}}$. The effect of F_{Ps} or F_R is to weight the integrands in equations (30) and (31) heavily towards small values of $\vec{\mathbf{Q}}$, $\vec{\mathbf{Q}}_a$, and $\vec{\mathbf{Q}}_b$ to the extent allowed by energy and momentum conservation. The primary advantage of writing H_{Ps} and H_R in terms of the helicity density is a simple unified semiquantitative understanding of the mechanism producing these asymmetries which is reminiscent of the qualitative "helical electron gas" model of Garay and Hrasko[9].

Equations (30) and (31) indicate that the signs of H_{Ps} and H_R are determined by the function $\hat{\mathbf{p}} \cdot \hat{\mathbf{q}} h_{Av}(q)$, since all of the other quantities appearing in these equations have a positive sign. In view of the expected difficulty of accurately calculating the sign of h_{Av} theoretically, an important question is whether the sign of H_R at high energies can be predicted from a knowledge of the sign of H_{Ps} at low energies, as the latter may be determined experimentally by the methods of ref. 7. As yet the dependence of h_{Av} on the bound-electron momentum q is not understood sufficiently to answer this question, but if it can eventually be answered in the affirmative, the positronium formation asymmetry measurements may then be able to provide the answer to the important question of the relative susceptibility of D compared with L isomers to ionization by the polarized electrons from nuclear β decay.

I thank G. W. Ford, D. W. Gidley, A. Rich, P. G. H. Sandars, J. Van House, and P. W. Zitzewitz for hospitality and for helpful and stimulating discussions.

Received 30 December 1981; accepted 7 April 1982.

1. Kizel, V. A. Soviet Phys. Usp. 23, 277–295 (1980).
2. Norden, B., J. molec. Evolut. 11, 313–332 (1978).
3. Keszthelyi, L. Origins of Life 8, 299–340 (1977); Origins of Life 11, nos. 1/2 (March/June 1981).
4. Rein, D. W. J. molec. Evolut. 4, 15–22 (1974).
5. Hegstrom, R. A., Rein, D. W. & Sandars, P. G. H. J. chem. Phys. 73, 2329–2341 (1980).
6. Vester, F., Ulbricht, T. L. V. & Krauss, H. Naturwisschaften. 46, 68 (1959).
7. Gidley, D. W., Rich, A., Van House, J. C. & Zitzewitz, P. W. Nature 297, 639–643 (1982).
8. Taylor, J. R. Scattering Theory (Wiley, New York, 1972).
9. Garay, A. S. & Hrasko, P. J. molec. Evolut. 6, 77–89 (1975).
10. Inokuti, M. Rev. Mod. Phys. 43, 297–347 (1971).
11. Zel'dovich, Ya. B. & Saakyan, D. B. Soviet Phys. JETP 51, 1118–1120 (1980).
12. Condon, E. U. Rev. Mod. Phys. 9, 432–457 (1937).
13. Mann, A. K. & Primakoff, H. Origins of Life 11, 255–265 (1981).

Secretory protein translocation across membranes—the role of the 'docking protein'

David I. Meyer, Elke Krause & Bernhard Dobberstein

Division of Cell Biology, European Molecular Biology Laboratory, D-6900 Heidelberg, FRG

On emergence of the signal sequence of nascent secretory proteins from the large ribosomal subunit, translation is stopped by a cytoplasmic protein complex. A specific membrane protein of the endoplasmic reticulum, the 'docking protein', releases this block and allows further synthesis to be directly coupled to transfer across the membrane.

NASCENT secretory proteins are extended at the N-terminus by 15–30 amino acids, the signal or leader sequence[1–3]. This extension, or its functional equivalent[4], is required for the vectorial transfer of the protein across the membrane of the endoplasmic reticulum (ER)[1,5–9]. On emergence of the signal sequence from the large ribosomal subunit, the ribosomal complex specifically makes contact with the membrane, then the nascent protein is translocated across the ER membrane. As the translocation phenomenon has been reconstituted *in vitro* using isolated rough microsomes[7,10], the molecules which mediate this transfer must be present in, and dissectable from, these membranes[11–16,17].

The ability to transport nascent secretory proteins is lost when rough microsomes are exposed to high salt concentrations[18]. A protein complex of molecular weight (MW) 250,000 (250K) is removed from the membrane in these conditions, which results in the loss of translocation activity[19,20]. Recently, it was determined that this 250K protein, referred to as the signal recognition protein (SRP), functions *in vitro* by selectively stopping the translation of nascent secretory proteins[21]. When salt-washed rough microsomes are added to the SRP-blocked system, translation continues and translocation of the nascent peptide occurs[21]. The question then arises as to which ER-specific membrane component is responsible for relieving the translation block. We have recently shown that a 60K peptide can be released from salt-washed rough microsomes by elastase and high salt treatment[15,16]. This component is ER-specific, is required for vectorial transfer of secretory proteins, and is derived from a membrane-bound protein of molecular weight 72K (ref. 22). So far, the role of the 72K protein has remained unknown.

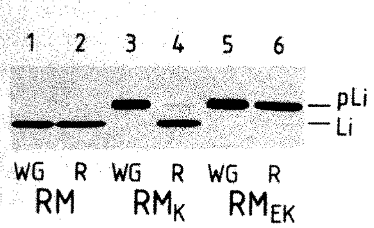


Fig. 1 Translocation activity of rough microsomes assayed in wheat germ and reticulocyte lysate. RM, whole rough microsomes; RM_K, KCl-washed RM; RM_{EK}, elastase-treated, KCl-washed RM; WG, wheat germ lysate; R, reticulocyte lysate system; pLi, pre-IgG light chain; Li, authentic IgG light chain. Membranes were prepared and *in vitro* translations were performed as described previously[15,18]. Translocation activity is measured as the ability of rough microsomes to convert pLi to Li when added to a cell-free system primed with mRNA encoding mainly IgG light chain. The figure shows a fluorogram depicting the translation products of *in vitro* synthesis.

Table 1 Ability of protein components to restore translocation activity to microsomal membranes

Microsomes/components	Activity in: Wheat germ	Reticulocyte lysate
RM	+	+
RM_K	−	+
RM_K + SRP	+	+
RM_K + DP_f	−	+
RM_{EK}	−	−
RM_{EK} + DP_f	−	+
RM_{EK} + SRP	−	−
RM_{EK} + DP_f + SRP	+	+

Rough microsomes (RM) in combination with isolated factors (DP_f or SRP) were tested for their ability to translocate nascent proteins in wheat germ and reticulocyte lysate systems. To exclude the possibility that reconstitution was being mediated by contaminating proteins in DP_f preparations, controls were done using a specific anti-DP_f antibody[22]. In these controls the antibody was used to specifically remove DP_f from the preparations. Such controls were performed in every case where DP_f was used in this report, and each time the preparation treated in this way was unable to reconstitute activity in inactive rough microsomes. RM_K, KCl-washed RM; RM_{EK}, elastase-treated and KCl-washed RM; SRP, signal recognition protein; DP_f, cytoplasmic fragment of docking protein released by elastase and high salt. Microsomes and factors were prepared and assays carried out as described previously[15,16,20].

We report here that this protein is indeed the molecule which relieves the SRP-induced block. Translation can then proceed when the initiated ribosomal complex has made contact with the correct membrane, that is, the one containing the 72K protein. In accordance with this function, we shall henceforth refer to this molecule as the 'docking protein' of the ER. We also report the presence in the cell cytoplasm of a functional equivalent of the microsomal signal recognition protein.

Signal recognition protein interacts with docking protein

Previously published data suggested that the source of the cell-free translation system affected greatly the translocation capability observed for salt-washed microsomes (compare refs 14, 15 with 19, 20). Whole rough microsomes (RM) were fully active in the wheat germ and reticulocyte lysate systems (Fig. 1, lanes 1 and 2). Salt-washed microsomes (RM_K), however, were only active in the reticulocyte lysate (compare Fig. 1, lane 3 with lane 4). Microsomes treated with protease and high salt (RM_{EK}) were inactive in either system (Fig. 1, lanes 5 and 6).

Salt washing exerts its effect by removing the 250K complex (SRP) from rough microsomes[19,20], whereas salt washing in conjunction with proteolysis removes from RM_K an active 60K cytoplasmic domain from the 72K docking protein (referred to as 'docking protein fragment', or DP_f)[16]. To define the requirements of translocation more precisely, we recombined the two factors with the various inactive rough microsomes and tested these in the two translation systems. The addition of the 60K fragment conferred activity on RM_{EK} in the reticulocyte lysate system alone; this combination is hence functionally equivalent to RM_K (see Table 1). It should then be possible to add SRP to $RM_{EK}+DP_f$ and restore activity in the wheat germ system. This was indeed the case, suggesting a functional identity between $RM_{EK}+DP_f+SRP$ and whole rough microsomes. We therefore conclude that two components are required to restore full activity to RM_{EK}: the 60K DP_f fragment and the 250K complex (SRP). It is clear also that reticulocyte lysate contributes the functional equivalent of the SRP.

The fact that the SRP and the DP_f could be removed sequentially from rough microsomes suggested a possible interaction between them. To test this, membranes (RM_{EK}) were incubated with either SRP or DP_f. After washing, the membranes were assayed in the wheat germ system in the presence of DP_f or SRP, respectively. In agreement with the results given above (Fig. 1), neither RM_{EK} (Fig. 2*a*, lane 1) nor RM_{EK} charged with DP_f (lane 2) were active. When the RM_{EK}–DP_f membranes were assayed in the presence of SRP (Fig. 2*a*, lane 3) the activity was restored to the level of whole rough microsomes (Fig. 2*a*, lane 5). No reconstitution was observed (Fig. 2*a*, lane 4) for RM_{EK} pre-incubated with SRP and assayed in the presence of DP_f. These data demonstrate that DP_f must be present on the membrane for SRP to interact functionally with rough microsomes.

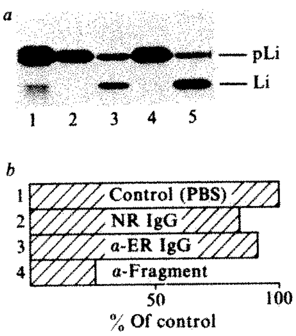


Fig. 2 Sequential reconstitution of translocation activity from isolated components. Inactive rough microsomes (RM_{EK}) were allowed to bind DP_f or SRP, and assayed in wheat germ in the presence of the component that was omitted from the preincubation. *a*, Lane 1, RM_{EK}; lane 2, RM_{EK} charged with DP_f; 3, RM_{EK} charged with DP_f assayed in the presence of SRP; 4, RM_{EK} charged with SRP assayed in the presence of DP_f; 5, intact rough microsome control. RM, DP_f and SRP were prepared as described elsewhere[15,16,20]. Activity is defined as described in Fig. 1 legend. *b*, RM_K were treated with anti-DP_f IgG, washed and assayed for translocation activity in the presence of SRP in a dithiothreitol-free wheat germ system. Lane 1, activity of RM_K pretreated with PBS; 2, activity of RM_K pretreated with normal rabbit (NR) IgG; 3, activity of RM_K pretreated with an affinity-purified anti-ER glycoprotein antibody specific for antigens on the surface of RM_K vesicles; 4, activity of RM_K pretreated with anti-DP_f IgG[22]. In the membrane-affinity-based translocation assay, 2.5 μl rough microsomes ($A_{280}=60\ ml^{-1}$) were incubated in the presence of the fraction being tested at 0 °C for 30 min in conditions where KCl concentration = 50 mM. The membranes were pelleted by centrifugation at 12,500 *g* for 20 min, then the rough microsome pellets were washed and suspended in the complete wheat germ cell-free translation system and assayed for translocation activity.

Further evidence for the functional interaction of DP_f with SRP is provided by data obtained using a specific anti-DP_f antibody[22]. When RM_K were charged with this antibody (by the method described in Fig. 2 legend), the membranes were no longer able to interact functionally with SRP (Fig. 2*b*). RM_K incubated with either pre-immune IgG or an ER-specific antibody directed against other antigens on the cytoplasmic surface of rough microsomes were still fully active when assayed in the wheat germ system in the presence of SRP.

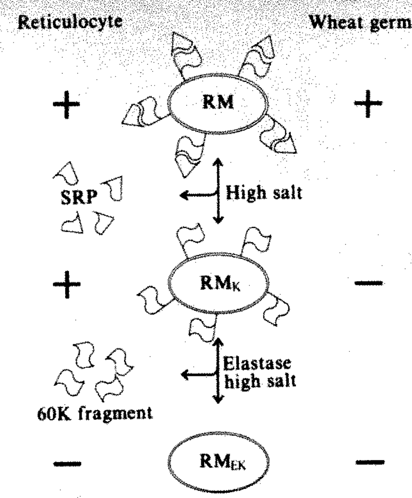


Fig. 3 Dissection and reconstitution of rough microsomes using protease and high salt. When rough microsomes are treated with 0.5 M KCl, translocation activity is lost in wheat germ (due to the salt-mediated removal of SRP) and RM_K are generated which are nonetheless active in reticulocyte lysate. Treatment of salt-washed microsomes (RM_K) with elastase and high salt yields RM_{EK} which are inactive in either system. This is due to the loss of a 60K peptide fragment from the docking protein (DP_f). Reconstitution can be achieved sequentially by re-addition of DP_f to RM_{EK} (thus restoring activity in reticulocyte lysate) followed by re-addition of SRP to RM_{EK}–DP_f membranes. In this way full reconstitution is achieved in both cell-free systems.

A model summarizing these findings is shown in Fig. 3. High salt concentration removes the 250K complex (SRP), yielding RM_K which are inactive in the wheat germ system. Proteolysis and high salt remove the 60K portion of a larger membrane protein (DP), yielding RM_{EK}, which is inactive in either system. Reconstitution occurs in the reticulocyte lysate when DP_f is added to RM_{EK} and in the wheat germ system when DP_f and then SRP are added to RM_{EK}.

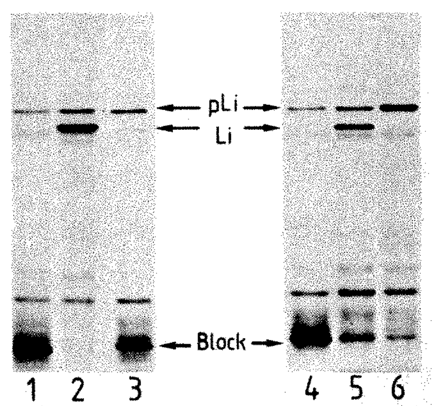


Fig. 4 SRP-induced block of translation of secretory proteins is removed by DP_f. Translations in the wheat germ system were blocked by the addition of SRP. We examined the ability of different membrane preparations to overcome this block. Lane 1, products of light chain mRNA-dependent cell-free translation (15 min) in wheat germ in the presence of SRP. Note the large accumulation of 8–9K material labelled 'block'. Lane 2, addition at 15 min of RM_K to a 'blocked' translation, observed at 40 min. 3, As in lane 2 except that RM_{EK} was added instead of RM_K. 4, Same as lane 1. 5, As in lane 2 except that $RM_{EK}+DP_f$ was added instead of RM_K. 6, As in lane 2 except that DP_f was added instead of RM_K (note the increase in pLi as opposed to Li due to the absence of membranes). Synchronous translation assays were carried out as described by Walter and Blobel[21].

DP fragment releases SRP-induced translation block

Walter *et al.*[21] have recently demonstrated that the SRP arrests the translation of preprolactin in the wheat germ cell-free system. This phenomenon is characterized by an accumulation of incomplete nascent peptides having a molecular weight of 7–8K. A similar result was obtained for IgG light chain (Fig. 4, lanes 1 and 4). This block is removed by the addition of RM_K (lane 2). Then the question arises: with which membrane component does SRP interact to result in the resumption of translation? DP_f, RM_{EK} or both could be required. As shown in Fig. 4, the block (lane 1) was removed by adding RM_K (lane 2), but not by addition of RM_{EK} (lane 3). When DP_f was added to the RM_{EK}, the block was removed (lane 5). The same effect was obtained when DP_f alone was added to the blocked system (lane 6). Thus we conclude that DP_f is necessary for the resumption of translation and translocation, while the residual membrane (RM_{EK}) is not.

Translation of secretory proteins blocked by soluble protein in reticulocyte lysate

The ability of RM_K to translocate nascent peptides in the reticulocyte lysate system, but not in wheat germ, suggested that a molecule equivalent to SRP must exist in the reticulocyte lysate. Such a component was characterized by making use of its ability to interact functionally with the membrane in low salt conditions. Lysate fractions were tested by allowing the component to bind to RM_K and assaying for reconstitution of translocation activity in the wheat germ system. The use of such an affinity-based assay was necessary, as direct addition of reticulocyte lysate inhibited the wheat germ cell-free system. Furthermore, it proved extremely sensitive. Several properties of the reticulocyte factor were determined; the results (Table 2) indicate that this factor has several of the basic features of SRP. It is an acidic protein of MW ~250,000, has a hydrophobic character, and blocks translation of nascent secretory proteins. The reticulocyte factor-induced block can be removed by adding RM_K, as was observed in the case of SRP. The presence of a block-inducing substance such as the reticulocyte factor in reticulocyte lysate may explain why the cytoplasmically-located nucleocapsid proteins of certain viral mRNAs were translated well and the membrane proteins only poorly[23]. Due to the small quantity of the reticulocyte factor in reticulocyte lysate, it has not been possible to determine whether its subunit composition is the same as that of SRP.

The cytoplasmic location of reticulocyte factor suggested that SRP may also be only transiently associated with the membrane, and that it is present in the cell as a soluble, cytoplasmic factor. Using the membrane-affinity assay described in Fig. 2 legend, we found that post-microsomal supernatants of dog pancreas contained a substance capable of restoring activity to RM_K when assayed in the wheat germ system. It possessed all the characteristics of reticulocyte factor (see Table 2) and SRP. In addition, a fivefold higher level of SRP-like activity was found in dog pancreas supernatant when compared with reticulocyte lysate (Fig. 5). These findings are consistent with the functional role of SRP in translocation. The delay in translation of secretory proteins, that is, until contact with the proper membrane has been made, would be optimally localized in the cytoplasmic compartment.

Table 2 Properties of the reticulocyte factor

Property	Method of determination
Proteinaceous	Proteolysis
Salt-linked to rough microsomes	Microassay
Acidic *p*I	DEAE binding*
Hydrophobic	ω-NH_2 pentyl agarose*
MW = 250,000	Sucrose gradient*
Blocks translation of secretory proteins	Translation in wheat germ†

The ability to reconstitute translocation activity of RM_K in the wheat germ system was tested by adding various lysate-derived fractions using the method described in Fig. 2 legend. For proteolysis, the reticulocyte lysate was treated with trypsin (20 μg ml^{-1}) for 20 min at 25 °C and then tested in the assay system described in Fig. 2 legend. In the microassay, the reticulocyte lysate was brought to 0.5 M in KCl and tested as above. * Ref. 20. † Ref. 21.

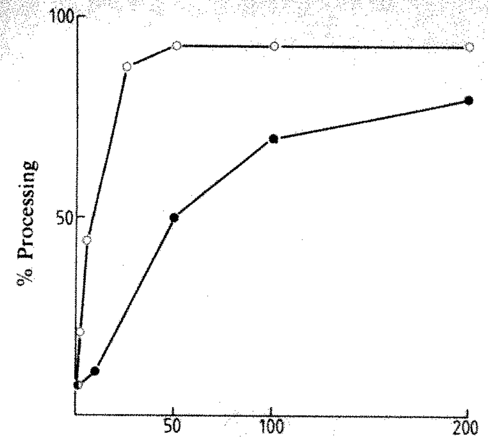


Fig. 5 SRP-like activity of reticulocyte lysate and dog pancreas post-microsomal supernatant. Reticulocyte lysate and dog pancreas supernatants (both depleted of ribosomes by centrifugation) were assayed for their ability to restore translocation activity to RM_K in the wheat germ system. 2.5 μl of RM_K (A_{280} = 60 ml^{-1}) were incubated with reticulocyte lysate or pancreatic soluble phase (both corrected to 40% (v/v) homogenates) in amounts indicated on the abscissa. The membranes were washed and assayed for translocation activity in wheat germ lysate. Activity is expressed as the % conversion of pLi to Li. ●, Reticulocyte lysate; ○, dog pancreas supernatant.

Implications and conclusions

A model consistent with all the data reported here, and with data obtained previously[14–16,18–22], is shown in Fig. 6. The initial events in protein translocation can be described as follows. Synthesis of a secretory protein is initiated on free ribosomes. After 60–70 amino acids have been polymerized and the signal sequence has emerged from the large ribosomal subunit, further translation is interrupted by SRP[21]. This block persists until contact is made with the 72K docking protein of which the 60K fragment (DP_f) represents the cytoplasmic domain. At this point translation, coupled with translocation, proceeds.

Such a sequence of events ensures optimal co-translational processing of secretory proteins, even when synthesis commences on free ribosomes. The presence of a cytoplasmic form of

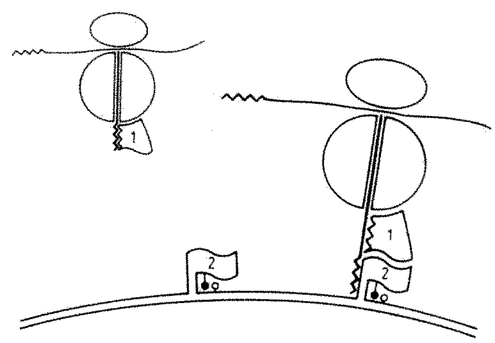


Fig. 6 Sequence of events in ER-specific translocation of secretory proteins. Initiation begins in the cytoplasm on free ribosomes. Translation is blocked by SRP (component 1) after 70–80 amino acids have been polymerized and the signal sequence emerges from the large ribosomal subunit (shown here in cross-section). This arrest of translation persists until contact is made with the 'docking' protein (component 2) which is a 72K, ER-specific membrane protein. Translation then resumes and translocation proceeds.

the signal recognition protein which blocks translation would guarantee that pre-secretory proteins are not completed in the cytoplasm. Efficient transfer of proteins can occur, therefore, only after the correct membrane has been contacted. Such a membrane can be operationally defined as one which contains the docking protein. This protein has been shown to be restricted to the endoplasmic reticulum[22].

The 72K docking protein is the first site of interaction between nascent secretory proteins and the ER membrane. An understanding of the function of this protein and SRP explains the specificity but not the mode of translocation. To determine the actual mechanism whereby proteins cross the membrane, further dissection of the membrane, including disassembly of the lipid bilayer, is necessary[17]. It is also conceivable that other soluble components provided by the cell-free translation systems function in this process. Their identification would require further fractionation of the lysates.

We thank John Stanger for photography and Wendy Moses for typing. This work was supported by a grant from the Deutsche Forschungsgemeinshaft.

Received 2 March; accepted 14 April 1982.

1. Milstein, C., Brownlee, G. G., Harrison, T. M. & Mathews, M. B. *Nature new Biol.* **239**, 117–120 (1972).
2. Kreil, G. A. *Rev. Biochem.* **50**, 317–348 (1981).
3. Sabatini, D. D., Kreibich, G., Morimoto, T. & Adesnik, M. *J. Cell Biol.* **92**, 1–22 (1982).
4. Lingappa, V. R., Lingappa, J. R. & Blobel, G. *Nature* **281**, 117–121 (1979).
5. Blobel, G. & Dobberstein, B. *J. Cell Biol.*, **67**, 835–851 (1975).
6. Blobel, G. & Sabatini, D. D. in *Biomembranes* (ed. Manson, L. A.) 193–195 (Plenum, New York, 1971).
7. Szczesna, E. & Boime, I. *Proc. natn. Acad. Sci. U.S.A.* **73**, 1179–1183 (1976).
8. Bedouelle, H. *et al. Nature* **285**, 78–81 (1980).
9. Talmadge, K., Stahl, S. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3369–3373 (1980).
10. Blobel, G. & Dobberstein, B. *J. Cell Biol.* **67**, 852–862 (1975).
11. Prehn, S., Tsamalonkas, A. & Rapoport, T. *Eur. J. Biochem.* **107**, 185–195 (1980).
12. Majzoub, J. A. *et al. J. biol. Chem.* **255**, 11478–11483 (1980).
13. Lingappa, V. R., Katz, F. N., Lodish, H. F. & Blobel, G. *J. biol. Chem.* **253**, 8667–8670 (1978).
14. Walter, P., Jackson, R. C., Marcus, M. M., Lingappa, V. R. & Blobel, G. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1795–1799 (1979).
15. Meyer, D. I. & Dobberstein, B. *J. Cell Biol.* **87**, 498–502 (1980).
16. Meyer, D. I. & Dobberstein, B. *J. Cell Biol.* **87**, 503–508 (1980).
17. Kreibich, G., Ulrich, B. L. & Sabatini, D. D. *J. Cell Biol.* **77**, 464–506 (1977).
18. Warren, G. & Dobberstein, B. *Nature* **273**, 569–571 (1978).
19. Dobberstein, B. *Hoppe-Seyler's Z. physiol. Chem.* **359**, 1469–1470 (1978).
20. Walter, P. & Blobel, G. *Proc. natn. Acad. Sci. U.S.A.* **77**, 7112–7116 (1980).
21. Walter, P. & Blobel, G. *J. Cell Biol.* **91**, 557–561 (1981).
22. Meyer, D. I., Louvard, D. & Dobberstein, B. *J. Cell Biol.* **92**, 579–583 (1982).
23. Revel, M. & Groner, Y. A. *Rev. Biochem.* **47**, 1079–1126 (1978).

Inducibility of human β-interferon gene in mouse L-cell clones

Hansjörg Hauser, Gerhard Gross, Wolfgang Bruns, Heinz-Kurt Hochkeppel, Ulrich Mayr & John Collins

Abteilung Genetik, Gesellschaft für Biotechnologische Forschung mbH, Mascheroder Weg1, D-3300 Braunschweig-Stöckheim, FRG

Transfer of a 36-kilobase piece of human DNA containing the β-interferon (IFN-β) gene into mouse Ltk$^-$ cells leads to transient expression of human interferon even without an exogenous inducer. A low level of human interferon expression is also found in most stable clones containing the transferred DNA. With double-stranded RNA or Newcastle disease virus (NDV) as inducer, human interferon expression is greatly increased. The induced transcript is identical to normal human IFN-β mRNA. Neighbouring genes contained on the transferred DNA are co-induced but are not essential for the production of human interferon in mouse L cells.

THE interferons (IFNs) are a class of proteins capable of inducing an antiviral state in many cell types against a wide range of viruses[1]. Almost all cell types produce one or more interferons upon induction with virus or a synthetic inducer such as double-stranded RNA[1]. FS4 cells, primary human fibroblasts, produce IFN-β (ref. 2) and several other proteins of unknown biological function[3] in response to the inducer poly(rI)poly(rC). In these cells interferon is normally synthesized for a limited period of several hours[4] during which the IFN-β mRNA accumulates and then rapidly disappears[5]. The shut-off occurs in the continuous presence of inducers and can be delayed by treatment with inhibitors of protein synthesis[6]. This treatment then allows the accumulation of IFN-β mRNA (superinduction)[5,7]. It has been suggested that continuous protein synthesis supplies the cell with IFN-β mRNA degrading protein(s), thus involving the participation of other genes in IFN-β regulation[6]. We have analysed the induction of human IFN-β and interferon-associated genes in human fibroblast and lymphoblastoid cells[8,9] and have previously reported the isolation and physical characterization of a 36-kilobase (Kb) region of the human genome including and surrounding an intact IFN-β gene[10]. The discovery that other poly(rI)poly(rC)-inducible genes are located in this cloned region presented the possibility that regulatory functions were acting through a common pathway on this family of genes.

We describe here the production of human interferon in mouse L-cell clones containing defined fragments from the human genome surrounding and including the human IFN-β (HuIFN-β) gene. The human IFN-β gene is clearly inducible with a spectrum of induction protocols which reflect the responsiveness of the mouse cells' own interferon genes rather than those of human fibroblasts. Because a 1.9-kb DNA fragment carrying only the structural gene of HuIFN-β is sufficient to allow induction, the neighbouring genes located on the human chromosomal DNA are not essential for the human interferon induction in mouse L cells.

Isolation of mouse L cells containing HuIFN-β cosmid DNA

We have co-transferred[11] cosmid pCosIFN-β DNA containing the human β-interferon gene[10] (Fig. 1) together with plasmid pHC79 2cos/tk DNA containing the herpes simplex virus thymidine kinase gene[12] into mouse Ltk$^-$ cells. More than 20 hypoxanthine–aminopterin–thymidine (HAT)-resistant clones were isolated. The single clones were maintained in culture for over 50 generations of HAT medium before further studies at the molecular level were carried out. During this period the properties of individual clones with respect to interferon production remained unaltered. High molecular weight DNA was prepared from these cell clones and compared with DNA from human diploid fibroblasts (FS4 cells), mouse Ltk$^-$ DNA and cosmid clone pCosIFN-β DNA by Southern blotting[13] after restriction enzyme digestion. Figure 2 shows the result for two

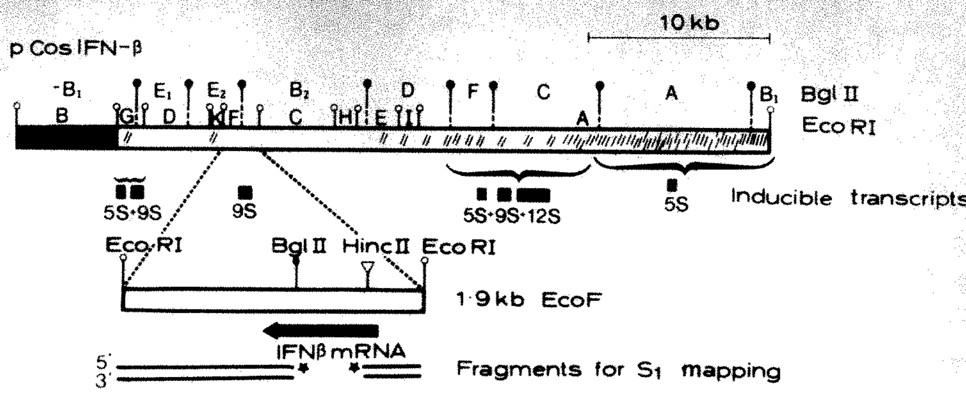


Fig. 1 Physical map of pCosIFN-β and transcription units inducible in FS4 cells using poly(rI)poly(rC) inducer. The upper bar shows a linearized form of pCosIFN-β including the vector pJB8 (filled bar)[10]. The human DNA part is shown as an open box containing the hatched regions which indicate sequences repeated on human chromosomal DNA. Transcripts hybridizing to pCosIFN-β sequences are depicted as black bars. The lower enlargement of the 1.9-kb *Eco* F fragment shows the location of the IFN-β gene transcript and the fragments (*Eco*RI–*Bgl*II and *Hin*cII–*Eco*RI) which were used for S_1-mapping of the interferon transcript (see Fig. 4).

cell clones. Hybridization with the 1.9-kb *Eco* fragment of the cosmid clone gives rise to a single band of the original size in FS4 DNA and DNA from both cell clones. Due to middle and highly repetitive DNA contained in the cosmid clone[10], hydridization of total cosmid to FS4 DNA leads to a background which obscures discrete bands.

Both mouse cell clones contain all *Eco* fragments present in the cosmid pCos IFN-β, indicating that most of the transferred DNA is in the original configuration. This could be explained by insertion of the transferred DNA into the genome in the form of concatenates. Such a tandem array would be a good substrate for cosmid rescue in which the *in vitro* packaging system is used to re-isolate the cosmid DNA directly out of the genome (according to ref. 11). When this experiment was carried out, pCosIFN-β clones could be re-isolated with an efficiency of ~20 clones per μg of mouse hybrid 2_4 genomic DNA (G.G., unpublished results), thus supporting the hypothesis that some, if not all, of the cosmid copies are integrated in tandem. By comparing the band intensities after different exposure times, we estimated that clone 12_3 has taken up >20 and clone 2_4 >50 copies of the total cosmid.

Active human interferon can be induced in mouse cell clones

All cell clones tested showed low levels of constitutive human interferon production (10–100 U ml^{-1}). On 4 h induction with poly(rI)poly(rC) plus DEAE–dextran (25 and 100 μg ml^{-1}, respectively), the interferon production in 19 out of 20 clones tested was enhanced to levels 4–1,000-fold higher than the constitutive base level production (Table 1). The levels of interferon production were determined by the antiviral assay with vesicular stomatitis virus as challenging virus on either FS4 or Vero cells (for assaying the human-specific activity) or mouse L cells (for assaying the mouse-specific activity)[1]. Although mouse L cells are known to produce α- and β-interferons on induction[14], tests on induced Ltk$^+$ control cultures for cross-reactivity showed no interferon activity in either human FS4 or monkey Vero cells. Furthermore, the antiviral activity of purified HuIFN-β was 1,000-fold higher on human than mouse test cells.

In clone 2_4, the Newcastle disease virus (NDV)-induced levels of HuIFN-β exceeded levels obtained with superinduced human FS4 fibroblasts, whereas mouse IFN production in the clone containing the HuIFN-β gene remained unaffected by the presence of the heterologous DNA. The human IFN-β gene responded to induction in the same way as the mouse genes: in conditions of superinduction, a very low level of interferon was detected and induction with NDV was more efficient than with other inducers (Table 1*a*). These data on the dominance of the mouse cell over the response pattern of the human interferon gene agree with the findings of Graves and Meager[15], who examined human interferon induction in human–mouse heterokaryotypes (from somatic cell fusions).

Wiranowski-Stewart *et al.*[16] have shown that the cell death during superinduction in human fibroblasts is largely due to the actinomycin D treatment and that if another RNA inhibitor is used, interferon induction may be repeated with the same cells at weekly intervals. We attempted repeated cycles of interferon induction with the mouse cells clone 2_4 using a 72-h cycle, and allowing 48 h for recovery in fresh medium. At least five cycles could be carried out without loss of interferon synthetic capacity (Table 1*b*).

Although there was little doubt that the interferon induced in the mouse cell clones and assayed on human cell lines was indeed HuIFN-β, we tested the characteristics of the protein further by immunological methods using monoclonal anti-HuIFN-β antibody[17]. The test involved gel electrophoresis of extracellular protein and electrophoretic blotting of this gel on

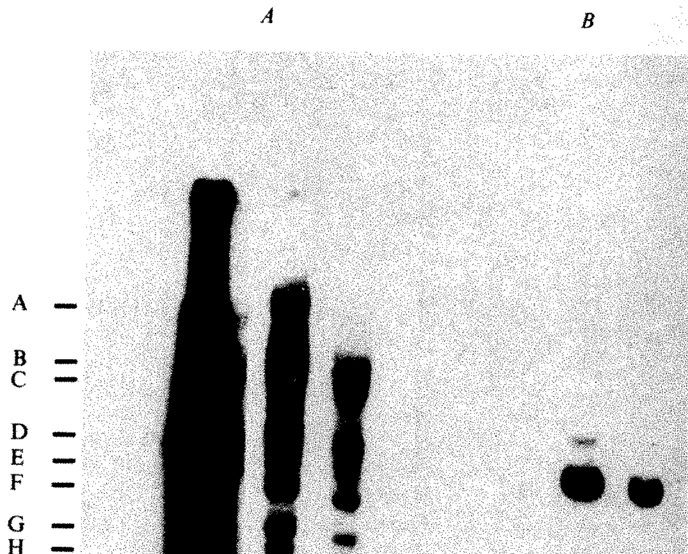


Fig. 2 Analysis of pCosIFN-β DNA in mouse L-cell clones. Mouse Ltk$^-$ cells were transformed using the calcium phosphate technique[17,32] with 250 μg supercoiled pHC79-2cos/tk$^+$DNA, 10 μg supercoiled pCosIFN-β DNA and 10 μg Ltk$^-$ carrier DNA per 5×10^5 cells.. Two established Ltk$^+$ clones (12_3 and 2_4) were analysed as follows. High molecular weight DNA from the clones was cleaved with *Eco*RI. 12μg of DNA was run in each lane of a 1% agarose gel, transferred (blotted) to nitrocellulose paper[13] and hybridized with either ^{32}P-labelled pCosIFN-β DNA(*A*) or *Eco* F fragment (*B*). *a*, Mouse Ltk$^-$; *b*, FS4 human fibroblast DNA; *c*, Ltk$^+$ clone 2_4; *d*, Ltk$^+$ clone 12_3. A–I indicate the size of *Eco* fragments typical of pCosIFN-β as in Fig. 1.

Table 1 Production of specific interferons in human and mouse cells

a Cell type	Interferon titre (U ml^{-1}) Without inducer	+poly(rI)poly(rC)	Superinduction	+NDV	Interferon type
Human FS4 fibroblasts	<5	1,600	12,000	2,000	Human
	<1	<1	<1	<1	Mouse
Mouse Ltk	<5	<5	<5	<5	Human
	<1	130	32	300	Mouse
Mouse Ltk$^+$ +pCosIFN-β clone 2_4	50	12,000	6,000	50,000	Human
	<1	130	32	300	Mouse
Mouse Ltk$^+$ +pCosIFN-β clone 12_3	10	3,200	ND	ND	Human
	<1	130			Mouse
b	Human interferon titre (U ml^{-1}) Day 1	Day 4	Day 7	Day 10	Day 13
Clone 2_4	6,500	3,200	12,000	6,500	6,500
Clone 12_3	3,200	1,600	3,200	1,600	ND

a, The supernatants were screened on FS4 or Vero cells, specifically for human IFN, and on mouse Ltk$^+$ cells, specifically for mouse IFN, in an antiviral test[1]. Human IFN activity is given as international reference units and mouse IFN in arbitrary laboratory units. Induction with DEAE-dextran plus poly(rI)poly(rC) (100 and 25 μg ml^{-1}, respectively) was performed for 4 h in serum-free Dulbecco's minimal essential medium (DMEM). For superinduction the cells were incubated for 4 h in serum-free medium with 25 μg ml^{-1} poly(rI)poly(rC) plus 50 μg ml^{-1} cycloheximide and 3.5–4 h after induction with 5 μg ml^{-1} actinomycin D. Induction with NDV Beaudette C was carried out by addition of 10 plaque-forming units of UV-inactivated NDV per cell in serum-free DMEM for 1 h. Following these induction periods the cells were washed thoroughly and incubated for a further 12 h in fresh medium plus 10% calf serum (10 ml per 2×10^7 cells) before assaying the IFN content in the medium. *b*, Confluent monolayers of L-cell clones 2_4 and 12_3 were induced for 4 h with DEAE-dextran-bound poly(rI)poly(rC) (see *a*). IFN produced after 12 h was in a minimal amount of serum-free medium (10 ml per 2×10^7 cells) after washing. The cells were allowed to recover for 48 h in medium plus serum and were then reinduced using the above protocol. ND, not determined.

to nitrocellulose paper ('Western blot'). Proteins having immunological cross-reaction with HuIFn-β were stained. As can be seen in Fig. 3, induced mouse Ltk$^-$ cells produced no cross-reacting material but the mouse clone 2_4 produced a major band identical in size to HuIFN-β (molecular weight 19, 500 (19.5 K)) as well as a minor band some 800 daltons smaller.

Transcription of human interferon gene in mouse cell hybrids

The transcription of the HuIFN-β gene in mouse cell clone 2_4 was studied by S_1 mapping[18,19] using labelled DNA fragments of the *Eco* F fragment (Fig. 1). The 5′ ends of the IFN-β mRNA from FS4 cells and mouse clone 2_4 were compared. The *Hinc–Eco* fragment encompassing the 5′ part of the transcribed and untranscribed human IFN-β gene was labelled at the *Hin*cII site and cut from the vector with *Eco*RI[9]. When hybrids between this fragment and total or poly(A)-containing RNA from FS4 and clone 2_4 cells were treated with S_1 nuclease and analysed on sequencing gels, a set of identical fragments of ~75 bases were protected against digestion by RNA from induced clone 2_4 cells and from superinduced FS4 cells, indicating that the inducible HuIFN-β transcripts are initiated at an identical site in both cell types (Fig. 4*A*). In contrast, no IFN-β RNA having distinct 5′ ends is present in uninduced cells of either type (Fig. 4*A*).

Nonspecific transcription starts can be deduced from the following experiment: the S_1 nuclease method was used with the *Bgl*II–*Eco* fragment comprising the 3′ part of the IFN-β gene region (Fig. 1) to detect transcripts undergoing specific termination within this region. RNAs terminating specifically within this sequence were present in uninduced as well as in induced clone 2_4 cells (Fig. 4*B*). This is in contrast to the transcription in FS4 cells in which no IFN-β gene transcripts are detectable in the non-induced cell.

To clarify the nature of these transcripts, we examined interferon-specific RNA by Northern blotting[20] in clone 2_4, identifying interferon-specific transcripts by hybridization with nick-translated *Eco* F DNA (Fig. 5). A 9S RNA species corresponding to IFN-β mRNA predominated over the background only after induction. The background present in RNA from induced and non-induced cells reflects transcripts of heterogeneous length mostly longer than 9S. The transcription of this heterogeneous RNA in mouse clone 2_4 is probably due to multiple non-specific transcription starts far upstream of the specific start site which is characterized by a 'TATA box' and

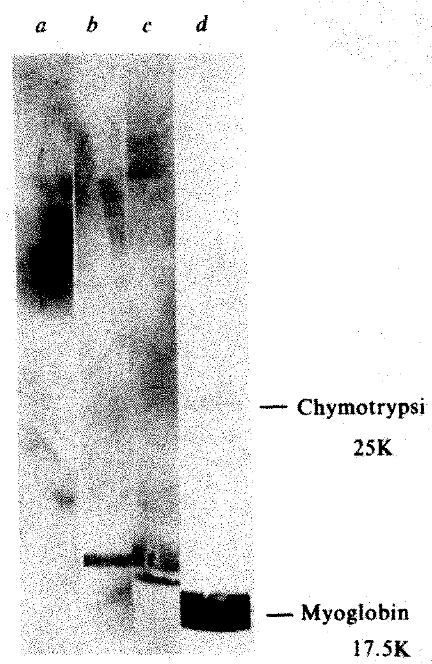
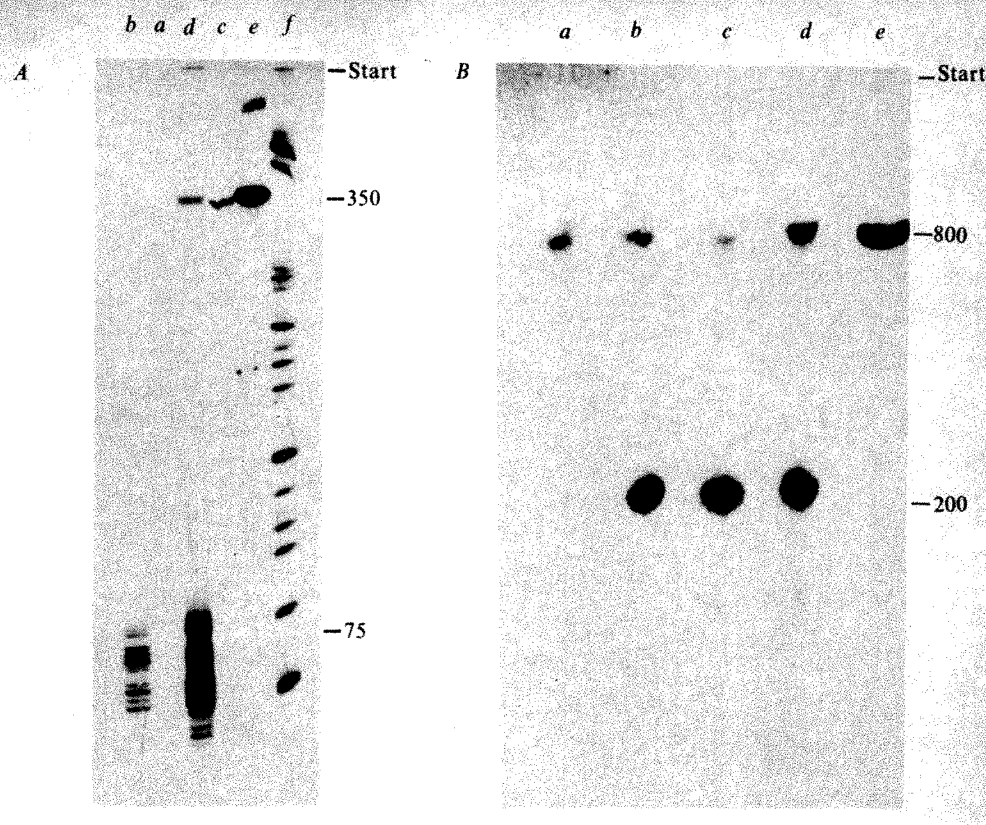


Fig. 3 Analysis of human IFN-β from a mouse cell hybrid using a monoclonal antibody-driven enzyme-linked immunosorbent assay. Supernatants were collected from mouse Ltk$^-$cells (*a*) or mouse Ltk$^+$ clone 2_4 cells (*c*) at 12 h after a 4-h induction period (poly(I)poly(C) (25μg ml^{-1})+ 100 μg ml^{-1} DEAE-dextran in serum-free DMEM). These supernatants were precipitated with 10% trichloroacetic acid and the pellets washed three times in acetone (0 °C) and once with cold ether. After lyophilization the pellets were suspended in SDS-sample buffer. As a positive control, human IFN-β from FS4 human fibroblasts was used in position *b*. Molecular weight markers were run in lane *d*. All these samples were run on a 15% SDS-polyacrylamide gel according to Laemmli[33]. This gel was blotted electrophoretically (a 'Western blot') onto a nitrocellulose sheet (Millipore) as described by Towbin *et al.*[34]. The section of the blot containing the molecular weight markers (*d*) was stained with amido black. The other sections were tested for human IFN-β content by the following treatment: (1) 2 h at 25 °C in 5% bovine serum albumin (BSA)/5% horse serum to saturate nonspecific protein binding sites; (2) wash in Tris-buffered saline, pH 7.4; (3) mouse monoclonal (ascites fluid: IgM) anti-human IFN-β was added at a 1:100 dilution in 5% BSA/5% horse serum/0.01% sodium azide and incubated for 24 h at 4 °C. After a further washing step (4) the blot was soaked for 16 h in rabbit (IgG) anti-mouse IgM (1:500 in 5% BSA/5% horse serum) and after a further wash (5) was incubated for 16 h with peroxidase-conjugated goat (IgG) anti-rabbit IgG (1:500 in 5% BSA/5% horse serum). The blot was washed again and the colour reaction performed with *o*-diasimidine as substrate. Lane *b* contained 30,000 U of human IFN-β and *c* contained 40,000 U of human interferon (as measured on human FS4 cells.)

Fig. 4 S_1 nuclease mapping of HuIFN-β mRNA made in human and mouse L-cell clones. Total RNA was extracted from FS4 cells 4 h after superinduction and from mouse clone 2_4 cells which had been induced for 7 h with DEAE–dextran-bound poly(rI)poly(rC). The ^{32}P-end-labelled probes (compare Fig. 1) for mapping the 5′ and 3′ ends from huIFN-β mRNA were prepared as described elsewhere[9]. 1 μg of poly(A)-containing RNA from FS4 cells and 25 μg total RNA from mouse cells were used for hybridization with an excess of the end-labelled DNA fragment in 80% formamide buffer. Hybridization was performed at 42° C for 12 h. The 5′ hybrids (*A*) were treated with a high concentration of S_1-nuclease (5,000 U ml^{-1}) to reduce the amount of unhybridized DNA fragment. For 3′-end mapping, 1,000 U ml^{-1} S_1 nuclease was used. The hybrids were denatured and run on a denaturing 6% acrylamide gel (*A* for the 5′ probe) or a 2% alkaline agarose gel (*B* for the 3′ probe), and the gels were dried and autoradiographed at −70 °C. *a*, Uninduced FS4 RNA; *b*, induced FS4 RNA; *c*, uninduced clone 2_4 RNA; *d*, induced clone 2_4 RNA; *e*, undigested DNA fragment used for hybridization; *f*, ^{32}P-end-labelled pBR322 DNA digested with *Hae*III; sizes are indicated in bases.



other sequences typical of eukaryotic promoters[10,21]. The protection of the complete *Hinc–Eco* fragment with RNA from induced and non-induced clone 2_4 cells (Fig. 4*A*, lanes *c*, *d*) reflects nonspecific transcripts which start 5′ upstream of the *Eco* site.

The presence of human IFN-β RNA with distinct 5′ ends in the induced mouse clone 2_4 is consistent with the production of HuIFN-β activity in this clone. Superinduced FS4 cells and induced mouse L cell hybrids express similar levels of HuIFN-β activity (Table 1*a*).

Transcription of pCosIFN-β DNA in mouse cell clones

In FS4 cells, in addition to IFN-β mRNA, several poly(rI)poly(rC) inducible transcripts can be detected which hybridize to the human DNA carried by pCosIFN-β: a small RNA of ~5S hybridizes to the *Bgl*II A, C and F fragments, a 9S RNA hybridizes to *Eco* G and A fragments and a transcript of ~12–13S hybridizes to fragment *Eco* A (ref. 8).

To determine whether the same pattern of RNAs is transcribed from the pCosIFN-β DNA in the mouse cell clones, we isolated total and poly(A)-containing RNA from hybrid cell clones before and at several times after induction. Transcripts from the transferred cosmid DNA were identified by hybridization to Northern blots of nick-translated DNA from pCosIFN-β or isolated DNA fragments. Heterogeneous transcripts were found corresponding to most parts of the transferred pCosIFN-β DNA, including the vector DNA (data not shown). In addition, specific bands were detected in the RNA from clone 2_4 after induction with poly(rI)(rC) plus DEAE–dextran (Fig. 5). The presence of the background of heterogeneous transcripts in the uninduced state prevented us from clearly identifying the large transcripts corresponding to *Eco* fragments A, G and F. Transcripts similar to the RNAs from FS4 cells can be seen as faint bands above the background (data not shown). Nevertheless, a small inducible transcript of ~5S (Fig. 5) which hybridises specifically to the *Bgl*II A, C and F fragments, could be clearly identified (data with individual probes not shown). This RNA is most probably identical to that seen in induced FS4 cells.

Regulation of human IFN-β expression in mouse L cells

To examine the influence of flanking sequences on the regulation of interferon gene expression, we co-transferred pBR13 carrying the 1.9-kb *Eco* F fragment of pCosIFN-β in pBR325 together with pHC79-2cos/tk into mouse Ltk$^-$ cells. Again, screening of single resistant cell clones for interferon production and induction showed a low constitutive base level production and variable inducibility of HuIFN-β (data not shown), indicating that the 1.9-kb *Eco* fragment is sufficient for interferon production and response to induction.

Comparison of the kinetics of human interferon induction and shut-off in cell clones containing either pCosIFN or pBR13 DNA showed no significant difference. This indicates that the neighbouring DNA regions present in pCosIFN-β are not necessary for the regulation of the human β-interferon gene in mouse L cells.

The biological assay provides a sensitive and specific tool for measuring even low amounts of interferon production. We have examined transient expression of pCosIFN-β and pBR13 DNA in mouse Ltk$^-$ cells and, using the calcium phosphate technique, found production of up to 100 U ml^{-1} of human interferon in the cell supernatant up to 3 days after transformation. However, there was no further induction by DEAE-bound poly(rI)poly(rC) during this period, although endogenous mouse interferon was inducible to a normal extent. This indicates that the ability of the IFN-β gene to respond to induction develops with the time after transfer.

Discussion

The availability of cloned genomic DNA and its introduction into heterologous mammalian cells allows the potential study of gene regulation in a foreign environment. Many attempts to find a gene which maintains its inducibility in a heterologous background have been unsuccessful (for example, chicken ovalbumin and rabbit β-globin genes in mouse cells[22,23]). However, recent findings with cloned mouse mammary tumour virus (MMTV) and rat α_2-microglobulin genes in mouse Ltk$^-$ cells have shown that hormone responsiveness is maintained in the new environment[24–26].

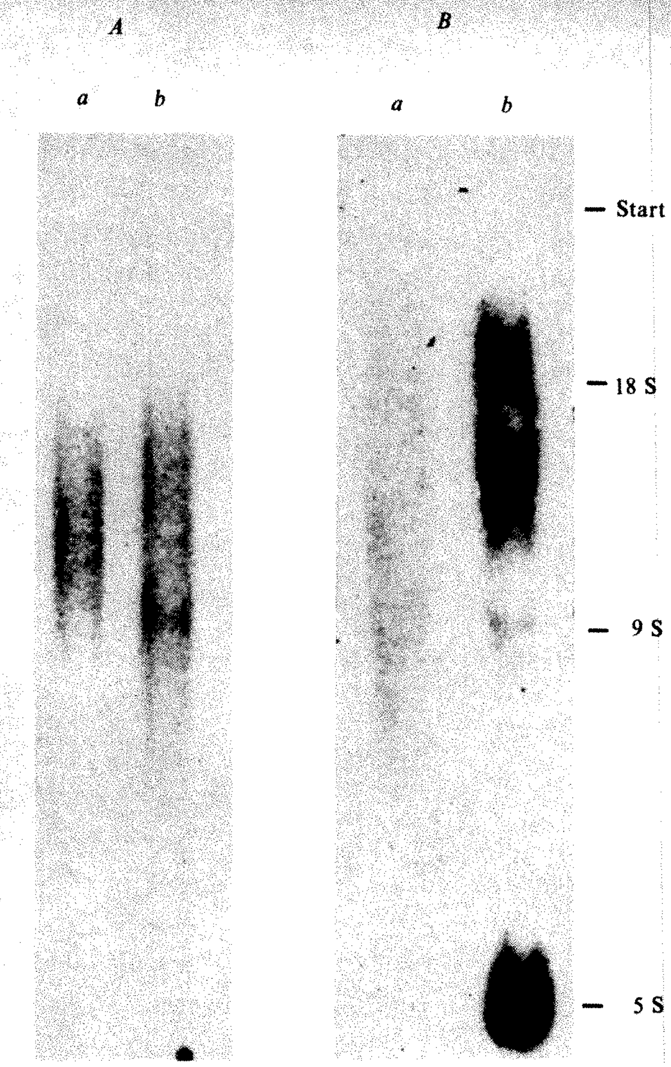


Fig. 5 RNA transcribed from pCosIFN-β in mouse Ltk$^+$ clone 2_4. Total and poly(A)-containing RNA from clone 2_4 cells before and at different times after induction with DEAE–dextran-bound poly(rI)poly(rC) were run on 2% formaldehyde gels[20]. The RNA was transferred to nitrocellulose filters and hybridized to nick-translated *Eco* F DNA (*A*) or pCosIFN-β (*B*)[8]. After washing, the dried filter was autoradiographed at $-70\,^{\circ}$C using an X-ray intensifying screen. Ethidium bromide-stained RNA served as size marker. *A*, 5 μg of poly(A)-containing RNA from uninduced (*a*) and 7 h-induced (*b*) cells. *B*, 20 μg of total RNA from uninduced (*a*) and from cells 12 h after induction (*b*). Nearly identical results with less intensive hybridization signals were obtained with RNA from clone 12_3 (not shown).

We have shown that in human fibroblasts and mouse L-cell clones containing the 36-kb piece of human DNA, the induced IFN-β mRNA is identical. However, a constitutive background of heterogeneous transcription starts is present in most mouse L-cell clones. A basal level of human interferon activity in the absence of inducer may be due to translation of IFN-β mRNA with heterogeneous size or, more probably, to a very low level of specific transcription, as is observed with chicken ovalbumin and rabbit β-globin genes in mouse L cells[22,23].

The human IFN-β produced, as measured by a species-specific test based on biologcial activity, was further defined immunologically with monoclonal antibody. Although the major protein species detected in the medium had a molecular weight identical to normally glycosylated HuIFN-β (molecular weight 19K) a smaller band was also observed. How these protein bands are actually related to normal human interferon remains an open question. The identity in size of the major band allows speculation that most of the HuIFN-β induced is glycosylated to the same extent as normal human IFN-β. (Mouse interferons have a larger molecular weight—24K and 36K (ref. 14).) The presence of the second smaller band could imply either a proteolytic cleavage or heterogeneity in the glycosylation.

We have previously shown that some of the transcribed regions adjacent to the huIFN-β gene are differentially regulated as, for example, in a comparison of human fibroblasts and lymphoblastoid cells[8]. Although these other transcripts are not as well defined as the IFN-β mRNA itself, it was clear that at least one other 'gene' (for the 5S RNA) was specifically inducible in the mouse clones to an even more dramatic extent than in human fibroblasts. As these neighbouring regions have little or no effect on the inducible transcription of HuIFN-β in mouse L cells, we conclude that if these transcribed regions are involved in the regulation of IFN-β expression in human fibroblasts, their function has been replaced by mouse gene products in the clones.

The finding that transient expression of the huIFN-β gene is unaffected by the presence or absence of inducer is an interesting phenomenon. It has been shown that transient expression of several genes occurs in a large proportion of transformed cells[27,28], although very few eventually stably maintain and express the gene. During stabilization of such clones the transferred DNA which is organized in a high molecular weight form (transgenomic form) becomes associated with chromosomes[29]. Most of the transiently expressed DNA is not in such a transgenomic organization[30]. The inducibility of the HuIFN-β gene correlates with its establishment in a stably integrated form.

During preparation of this manuscript, Ohno and Taneguchi[31] reported inducible transcription of RNA complementary to the *Eco* F fragment in mouse cells transformed with SV2 *Eco* gpt *Eco* F DNA.

We thank Dr Rentschler AG (now Bioferon AG) for providing us with induced FS4 cells and human β-interferon standards, Karin Littmann for technical assistance, Dr A. C. R. Samson for providing us with a range of NDV subtypes and Dr W. Lindenmaier for helpful discussion. A preliminary report of this work was presented at the 17th Harden Conference on Interferon, Ashford (September, 1981).

Note added in proof: Mantei and Weissmann[35] have demonstrated inducibility of human IFN-α1 mRNA in NDV-induced mouse L cell clones.

Received 1 March; accepted 7 May 1982.

1. Stewart, W. E. II *The Interferon System* (Springer, Vienna, 1979).
2. Field, A. K., Tytell, A. A., Lampson, G. P. & Hillemann, M. R. *Proc. natn. Acad. Sci. U.S.A.* **58**, 1004–1010 (1967).
3. Raj, N. B. K. & Pitha, P. M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4918–4922 (1980).
4. Vilcek, J., Ng, M. H. & Friedman-Kien, A. E. *J. Virol.* **2**, 648–650 (1968).
5. Raj, N. B. K. & Pitha, P. M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 7426–7430 (1981).
6. Vilcek, J., Havel, E. A. & Kohase, M. *J. infect. Dis.* **133**, A22–A29 (1976).
7. Tomkins, G. M. *et al. Cold Spring Harb. Symp. quant. Biol.* **31**, 349–360 (1966).
8. Gross, G., Mayr, U. & Collins, J. *The Biology of the Interferon System* (eds De Maeyer, E., Galasso, G. & Schellekens, H.) 85–90 (Elsevier, Amsterdam, 1981).
9. Gross, G., Bruns, W., Mayr, U. & Collins, J. *FEBS Lett.* **139**, 201–204 (1982).
10. Gross, G. *et al. Nucleic Acids Res.* **9**, 2495–2507 (1981).
11. Wigler, M. *et al. Cell* **16**, 777–785 (1979).
12. Lindenmaier, W., Hauser, H. Wilke, J. & Schütz, G. *Nucleic Acids Res.* **10**, 1243–1256 (1982).
13. Southern, E. *Meth. Enzym.* **68**, 152–176 (1980).
14. Fujisawa, J., Iwakura, Y. & Kawade, Y. *J. biol. Chem.* **253**, 8677–8679 (1978).
15. Graves, H. E. & Maeger, A. *J. gen. Virol.* **47**, 489–495 (1980).
16. Wiranowska-Stewart, M., Chudzio, T. & Stewart, W. E. *J. gen. Virol.* **37**, 221–223 (1977).
17. Hochkeppel, H. K., Menge, U. & Collins, J. *Eur. J. Biochem.* **118**, 437–442 (1981).
18. Berk, A. J. & Sharp, P. A. *Cell* **12**, 721–732 (1977).
19. Weaver, R. F. & Weissmann, C., *Nucleic Acids Res.* **7**, 1175–1193 (1979).
20. Thomas, P. S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201–5205 (1980).
21. Degrave, W., Derynck, R., Tavernier, J., Haegeman, G. & Fiers, W. *Gene* **14**, 137–143 (1981).
22. Breathnach, R., Mantei, N. & Chambon, P. *Proc. natn. Acad. Sci. U.S.A.* **77**, 740–744 (1980).
23. Wold, B. *et al. Proc. natn. Acad. Sci. U.S.A.* **76**, 5684–5688 (1979).
24. Hynes, N. E., Kennedy, N., Rahmsdorf, U. & Groner, B. *Proc. natn. Acad. Sci., U.S.A.* **78**, 2038–2042 (1981).
25. Lee, F., Mulligan, R., Berg, P. & Ringold, G. *Nature* **294**, 228–232 (1981).
26. Kurtz, D. T. *Nature* **291**, 629–631 (1981).
27. Capecchi, M. R. *Cell* **22**, 479–488 (1980).
28. Banerji, S. R. & Schaffner, W. *Cell* **27**, 299–308 (1981).
29. Scangos, G. & Ruddle, F. H. *Gene* **14**, 1–10 (1981).
30. Rusconi, S. & Schaffner, W. *Proc. natn. Acad. Sci. U.S.A.* **78**, 5051–5055 (1981).
31. Ohno, S. & Taneguchi, T. *Nucleic Acids Res.* **10**, 967–977 (1982).
32. Graham, F. L. & Van der Eb, A. J. *Virology* **52**, 456–467 (1973).
33. Lämmli, U. K. *Nature* **227**, 680–685 (1970).
34. Towbin, H., Steahelin, T. & Gordon, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4350–4354 (1979).
35. Mantei, N. & Weissmann, C. *Nature* **297**, 128–132 (1982).

Sequence homology and structural comparison between the chromosomal human α_1-antitrypsin and chicken ovalbumin genes

Margaret Leicht, George L. Long*, T. Chandra, Kotoku Kurachi*, Vincent J. Kidd, Myles Mace Jr, Earl W. Davie* & Savio L. C. Woo†

Howard Hughes Medical Institute Laboratory, Department of Cell Biology, Baylor College of Medicine, Houston, Texas 77030, USA
* Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA

The human chromosomal α_1-antitrypsin gene has been cloned. This gene is approximately 5 kilobase pairs long and contains three intervening sequences in the peptide-coding region. DNA sequences coding for the amino and carboxyl termini of α_1-antitrypsin have been identified. Human α_1-antitrypsin and chicken ovalbumin show significant sequence homology and belong to a common protein super-family. Yet the number, position and size of intervening sequences reveal that the two genes are dissimilar.

α_1-ANTITRYPSIN is an important plasma protease inhibitor which is capable of inhibiting a wide variety of serine proteases, including proteases involved in blood coagulation, fibrinolysis and kinin generation[1-4]. Although α_1-antitrypsin is a plasma protein of hepatic origin, its primary site of physiological action may not be in the blood. There is evidence that it is transported by passive diffusion into the alveolar structure of the lung and protects this organ from destruction by polymorphonuclear leukocyte elastase[5-9]. α_1-Antitrypsin inhibits elastase by forming an equimolar complex of enzyme and inhibitor with an association rate constant of 6.5×10^7 mol^{-1} (refs 10, 11). This association rate constant is 3.4×10^5 and 1.3×10^6 times greater than that of human plasmin and thrombin respectively. The balance between elastase and α_1-antitrypsin in the lung is perturbed genetically in individuals with an inborn deficiency of α_1-antitrypsin. This imbalance can lead to hydrolysis of the elastic fibres of the lung by elastase, resulting in permanent damage to the alveolar structure[12-14]. Accordingly, individuals with α_1-antitrypsin deficiency are 20–30 times more likely to develop chronic obstructive pulmonary emphysema than is the general population[15,16].

The synthesis of α_1-antitrypsin is controlled by an autosomal and allelic system[4,17], and over 30 different phenotypes of the protein have been identified[18,19]. In order to study the normal and abnormal α_1-antitrypsin genes, we have screened a human genomic DNA library using a cDNA probe prepared from baboon liver[20,21]. We report here the isolation and characterization of several clones containing the human chromosomal α_1-antitrypsin gene.

Overlapping genomic α_1-antitrypsin clones

Sixteen independent phage isolates were obtained when 2×10^6 plaques from the human genomic DNA library[22] were screened using the baboon α_1-antitrypsin cDNA clone[20] as hybridization probe. Subsequent analysis of the isolates indicated that they originated from four independent clones—αAT135, αAT35, αAT80 and αAT101. The four clones were analysed by restriction mapping and Southern hybridization using as probes a *Mbo*II fragment of pBaα1a1 DNA, which contains the 3'-terminal region of the baboon cDNA[20], and a *Hha*I fragment of pBAα1a2 DNA which is a baboon cDNA clone lacking only ~100 nucleotides at the 5' end of the mRNA[21]. These results have established the orientation of the human α_1-antitrypsin gene and indicate that the entire gene may reside within a 9.6 kilobase (kb) *Eco*RI DNA fragment in the human genome (Fig. 1).

† To whom correspondence should be addressed.

Mosaic structure of the human α_1-antitrypsin gene

The overall structure of the human α_1-antitrypsin gene was established by electron microscopy of hybrid molecules formed between the cloned chromosomal DNA and baboon α_1-antitrypsin mRNA. The mature mRNA consists of ~1,400 nucleotides. DNA was denatured thermally and hybrids were formed subsequently in conditions favouring RNA–DNA hybridization but not DNA–DNA reassociation. Figure 2*a* shows two representatives of 15 such molecules and the corresponding line

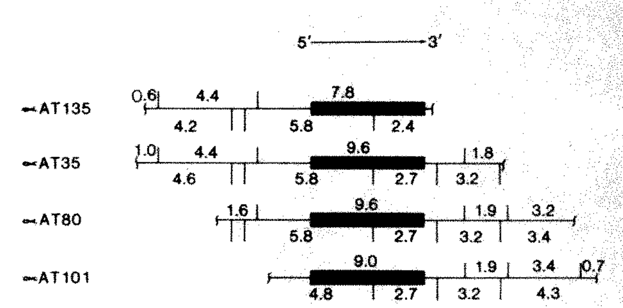


Fig. 1 Schematic representation of overlapping genomic clones of the human chromosomal α_1-antitrypsin gene. ■, The α_1-antitrypsin gene region; →, transcription orientation. The vertical lines above and below the horizontal lines represent *Eco*RI and *Hin*dIII cleavage sites respectively. The numbers indicate the distances between restriction sites in kilobases. The signs at the end of each clone represent the artificial *Eco*RI sites created during construction of the genomic DNA library by the method of Lawn *et al.*[22]. Approximately 2×10^6 recombinant Charon 4A phages were screened for clones containing α_1-antitrypsin gene sequences by the *in situ* plaque hybridization technique of Benton and Davis[56], using an amplification procedure previously described[57]. A hybridization probe for the α_1-antitrypsin gene was prepared by *Pst*I and *Mbo*II digestion of a partial-length cDNA clone containing the 3'-terminal sequence of baboon α_1-antitrypsin mRNA[20]. The resulting 120 bp fragment was purified by preparative polyacrylamide gel electrophoresis[58] and the DNA fragments were labeled with ^{32}P by nick translation according to the procedure of Maniatis *et al.*[59]. As the probe was only 120 bp long and of baboon origin, the library screening by cross-hybridization was carried out in slightly relaxed conditions (6×SSC at 62 °C). The α_1-antitrypsin gene-containing recombinant phages were plaque-purified three times and phage DNA was prepared as previously reported[60]. The genomic clones of the human α_1-antitrypsin gene were digested with a variety of restriction enzymes in conditions suggested by the manufacturer, and the resulting DNA fragments were separated by agarose gel electrophoresis. DNA fragments were transferred from the agarose gel to nitrocellulose filters by the method of Southern[61] and the filters were treated according to the procedure of Denhardt[62]. Hybridization was carried out overnight with the labelled baboon cDNA probes in 6×SSC at 68 °C. Differential hybridizations involving either the 5' or the 3' portion of a baboon cDNA probe were performed as described above after bidirectional transfer of DNA fragments from agar gels onto nitrocellulose filters by the method of Smith and Summers[63].

0028-0836/82/250655-05$01.00

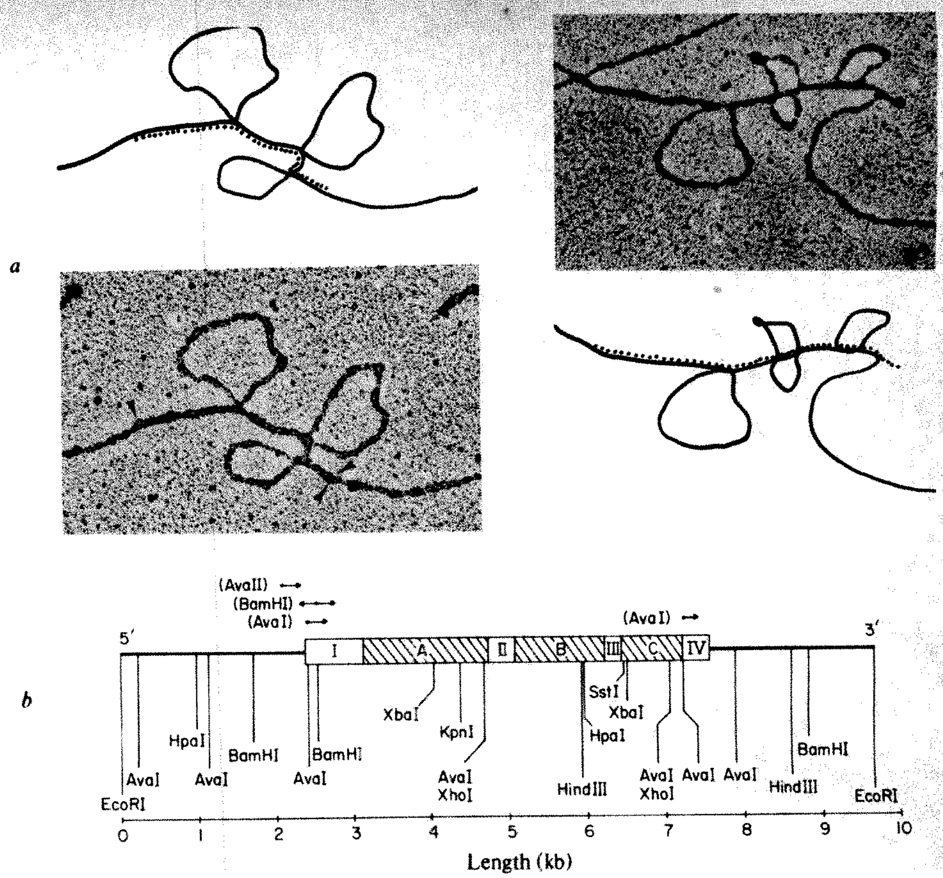


Fig. 2 Structure of the human chromosomal α_1-antitrypsin gene. *a*, Electron micrographs and line drawings of hybrid molecules of the human chromosomal α_1-antitrypsin gene and baboon liver poly(A)-RNA. ——, Single-stranded genomic human DNA; · · · · ·, baboon α_1-antitrypsin mRNA. Hybrids of the cloned human chromosomal α_1-antitrypsin gene (10 μg ml^{-1}) with the baboon liver poly(A)-RNA (20 μg ml^{-1}) were formed in 70% deionized formamide containing 100 mM Tris-HCl (*p*H 7.6), 10 mM Na^+-EDTA and 30 mM NaCl. The mixture was heated at 80 °C for 5 min and incubated at 55 °C for 3 h. Immediately after incubation, samples were prepared for electron microscopy by diluting 0.1–0.5 μg of nucleic acid with 100 μl of a solution containing 70% formamide, 0.1 M Tris-HCl (*p*H 8.4), 0.01 M Na^+-EDTA and 100 μg ml^{-1} of cytochrome *c*. The mixture was spread onto a hypophase of distilled water. Samples were picked up on collodion-coated 300-mesh copper grids, stained with uranyl acetate, rotary-shadowed with platinum-palladium and examined at 80 kV using a JEOL 100 C electron microscope. Molecules were measured using a Numonics electronic planimeter with ϕX174 RD DNA as the internal standard. *b*, Schematic representation of the human chromosomal α_1-antitrypsin gene. Structural segments (☐) are indicated by Roman numerals and intervening sequences (▨) by capital letters. The transcriptional orientation of the α_1-antitrypsin gene in human genomic DNA is from left to right. All restriction digests were performed in conditions suggested by the manufacturer. Horizontal arrows above the gene represent sequenced DNA regions (see Fig. 3). In each case, the arrow shows the direction of sequencing and the restriction enzyme site (in parentheses) that was labelled (closed circle).

drawings. It is evident that there are three intervening DNA loops (introns) of various sizes in the human α_1-antitrypsin gene. The poly(A) tract in the mRNA is clearly visible in the hybrid molecule shown in the right-hand panels, thereby confirming the orientation of the gene. Indeed, when αAT135 DNA was cleaved with *Eco*RI before hybrid formation with the baboon mRNA, the smallest intervening DNA loop was very close to one end of the DNA molecule (not shown). Numonic measurements of the hybrid molecules have indicated that the exon regions I, II, III and IV are approximately 0.71, 0.33, 0.13 and 0.27 kb long respectively. Introns A, B and C are 1.45, 1.15 and 0.8 kb long respectively, and all of them appear to be located in the 3′ half of the mRNA.

To characterize further the human chromosomal α_1-antitrypsin gene, the 9.6 kb *Eco*RI DNA fragment was subcloned into an *Eco*RI site of pBR322. The resulting clone, pAT9.6, was analysed by restriction mapping and Southern hybridization. Four exon segments were identified in the 9.6 kb *Eco*RI DNA fragment using a combination of enzymes that do not cut the baboon α_1-antitrypsin cDNA insert in pBaα1a2 (ref. 21). These results (Fig. 2*b*) confirmed the existence of three introns in the human α_1-antitrypsin gene. As the baboon cDNA clone was used to map the human chromosomal gene, the restriction enzymes that did not cleave the baboon cDNA might actually have cleaved the exons in the human gene due to sequence divergence between the two species. This possibility seemed unlikely, however, because at least two hexanucleotide recognition sites are present in each of the three introns in the human chromosomal DNA and absent from the baboon cDNA sequence. The presence of only three introns in the peptide-coding region of the human chromosomal α_1-antitrypsin gene has recently been confirmed by DNA sequence analysis (G.L.L., unpublished results).

The 5′- and 3′-terminal sequences of the human α_1-antitrypsin gene

Southern hybridization analysis between different portions of the baboon cDNA clone and human genomic fragments generated by digestion of pAT9.6 revealed DNA fragments which hybridize uniquely with either the 5′ or the 3′ end of the cDNA probe. By DNA sequencing, we identified fragments of the human genomic DNA that code for amino acids at both the amino- and carboxyl-terminal regions of human α_1-antitrypsin (Fig. 3). The two regions are ~5 kb apart, which is in good agreement with our estimates of the size of the gene, based on the electron micrographs. The amino acid sequence at the amino-terminal region is consistent for 30 of the 33 residues with that previously published for human α_1-antitrypsin[23]. Different amino acids include Lys 10, His 20 and Ile 26, which were reported as Glu, Ser and Leu respectively[23]. The DNA sequence corresponding to the amino-terminal region of the protein was confirmed by sequencing both strands of the genomic DNA. Furthermore, the residues in question are identical to those determined for baboon α_1-antitrypsin[21]. The amino acid sequence containing 32 residues at the carboxyl end of the mature protein was also deduced from the genomic DNA sequence. This amino acid sequence is in complete agreement with that previously published for α_1-antitrypsin[24]. Also, the genomic DNA sequence was identical with the corresponding nucleotide sequence of a human α_1-antitrypsin cDNA clone[21].

The first ATG start codon at the 5′ end of the α_1-antitrypsin gene is located 24 amino acids upstream from the amino-terminal Glu residue in the mature protein (Fig. 3*a*). This region seems to code for a typical signal peptide, which is removed from the mature protein during intracellular processing, before extracellular transport. This signal peptide resembles other

signal peptides in having an amino-terminal methionine residue, a hydrophobic core flanked by regions of more polar residues, a small uncharged amino acid at the putative cleavage site, proline at position −5 and a length of ~15–30 amino acids[25]. Furthermore, there appears to be a 'TATA box'[26] at position −25 to −31 of the gene (Fig. 3a), which resembles the consensus sequence, TATA$^{TAT}_{ATA}$, proposed by Cordon et al.[27]. The transcription start point in eukaryotes also has a consensus sequence, PyCAPyPyPyPyPy (A = position +1; Py = pyrimidine), based on 20 genes[27]. This is less obvious in the α_1-antitrypsin gene but can be tentatively assigned to bases −2 to 6 (Fig. 3a). Benoist et al.[28] have pointed out an additional consensus sequence (GG$^{C}_{T}$CAATCT) further upstream (~80 residues) from the transcription start point which may also be important in the control of gene expression. This 'CAAT' box is less recognizable in the α_1-antitrypsin gene.

The hexanucleotide AAUAAA occurs in eukaryotic mRNAs approximately 20 bases upstream from the point of poly(A) attachment[29]. However, in the α_1-antitrypsin gene, a sequence of AUUAAA is present in this position (Fig. 3b). This somewhat atypical sequence has also been reported for mouse pancreatic α-amylase[30], human leukocyte interferon[31] and chicken lysozyme[32] mRNAs.

Sequence homology between human α_1-antitrypsin and chicken ovalbumin

An unexpectedly high degree of amino acid sequence homology between the carboxyl-terminal 152 residues of human α_1-antitrypsin and those of chicken ovalbumin was originally observed by Hunt and Dayoff[33]. Based on the assumption that the rates of change for both proteins lie between 9.8 and 18 PAMs (point mutations per 100 residues) per 100 million years, which are accepted values for animal lysozyme and for pancreatic secretory trypsin inhibitor[34], it was proposed that the divergence time between the two proteins could be over 500 Myr ago. They then proposed that the two proteins belong to a common super-family and that they diverged before or during early vertebrate evolution[33]. We have subsequently determined the complete amino acid sequence of baboon α_1-antitrypsin and shown that the two seemingly unrelated proteins have a 24% amino acid sequence homology[21]. The amino acid sequences of the two proteins were also examined by a matrix-plot computer analysis[35,36], which reveals not only the extent of sequence homology, but also the positions of homologies, and relatedness of two sequences is displayed as a diagonal line on the plot. The diagonal line apparent in such a plot between α_1-antitrypsin and ovalbumin has confirmed that homology between the two proteins exists at corresponding positions throughout the molecules (data not shown).

Structural comparison of human α_1-antitrypsin and chicken ovalbumin genes

Gilbert[37] has proposed that intervening sequences may serve to separate portions of the coding region of genes corresponding to the structural–functional domains for their corresponding proteins. The gene structures for chicken ovomucoid[38], the β-chain of haemoglobin[39] and the heavy chain of immunoglobulin[40] support this hypothesis. As human α_1-antitrypsin and chicken ovalbumin apparently belong to the same super-family, it was predicted that the two genes would contain similar exon–intron patterns[33]. The number and locations of the introns in the human α_1-antitrypsin gene as determined by electron microscopic and restriction mapping were thus compared with those in the chicken ovalbumin gene (Fig. 4). The chicken ovalbumin gene contains seven introns and their exact locations in the gene have been established by DNA sequencing[41–43]. The human α_1-antitrypsin gene contains only three introns, which are located in the 3′ half of the corresponding mRNA

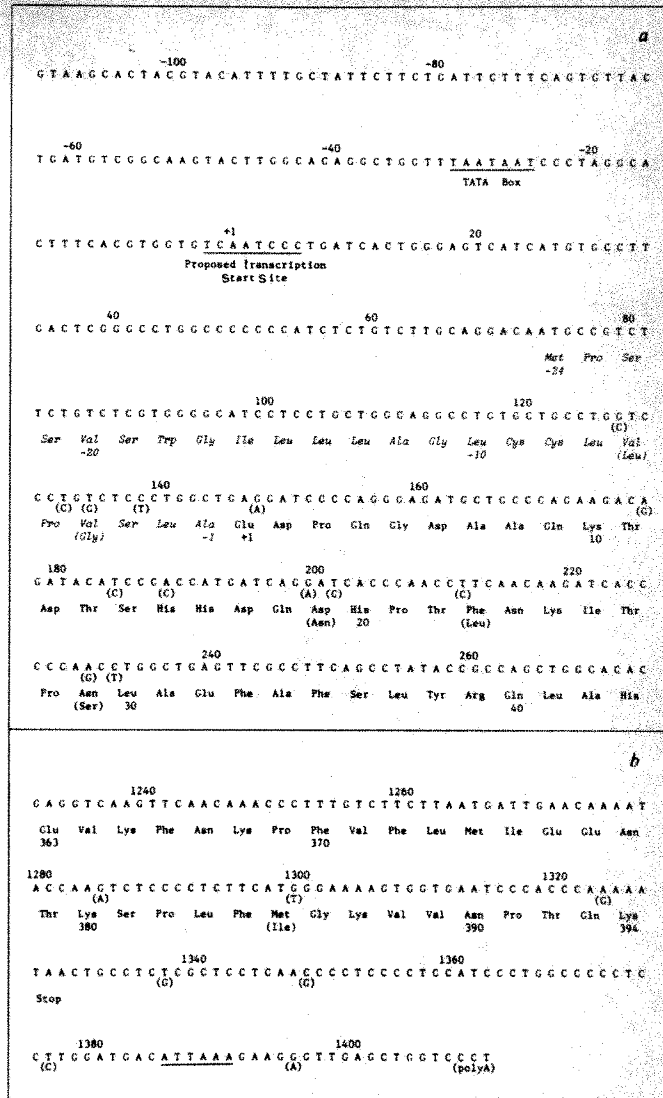


Fig. 3 Nucleotide sequences and the deduced amino acid sequences at the 5′ (a) and 3′ (b) termini of the human α_1-antitrypsin gene. Also included are the corresponding bases or amino acids (shown in parentheses) observed for baboon cDNA[21]. Numbering for the nucleic acids is based on the proposed transcription start point (+1), that of the amino acids on the amino terminal residue of the mature protein (+1). Proposed transcription recognition sites are underlined. The treatment of the DNA fragments with bacterial alkaline phosphatase, labelling of the 5′ termini with [γ-^{32}P]ATP using T4 polynucleotide kinase, isolation of labelled DNA fragments by preparative polyacrylamide gel electrophoresis and chemical degradation of end-labelled DNA fragments were performed according to the procedure of Maxam and Gilbert[58]. Numbering for the nucleic acids at the 3′ end and the corresponding amino acids is based on the transcription start point shown in a and homology with a baboon cDNA clone[21]. The underlined sequence corresponds to a site related to mRNA polyadenylation described by Proudfoot and Brownlee[29].

(Fig. 2), whereas all seven introns in the chicken ovalbumin gene are located in the 5′ half of the mRNA. Based on the electron microscopic map, the only introns that might share a common site in the two genes are intron A in the human α_1-antitrypsin gene and intron E or F in the chicken ovalbumin gene (Fig. 4). Intron A in the human α_1-antitrypsin gene has recently been located by DNA sequencing at a position 27 and 21 amino acids away from ovalbumin introns E and F respectively (Fig. 5). In the region shown, 21 of 60 amino acids (35%) and 86 of 180 nucleotides (48%) are identical. However, a comparison of nucleotide sequences for the 5′ and 3′ ends of the introns with the corresponding coding regions of the other gene shows only 28% (α_1-antitrypsin intron A) and 25% (ovalbumin intron F) identity of bases, values close to the 25% level

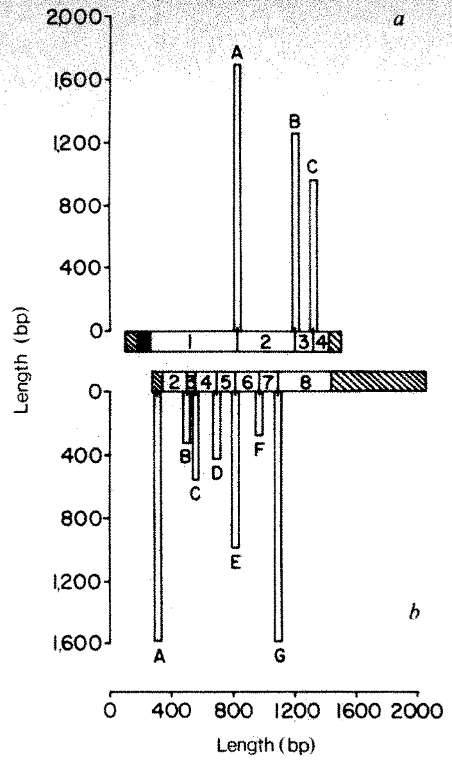


Fig. 4 Comparison of the genes for α_1-antitrypsin (a) and ovalbumin (b). □, Coding regions for the two proteins; ■, position of a proposed 'signal' sequence in α_1-antitrypsin; ▧, flanking noncoding regions of the corresponding mRNA transcripts. Alignment of the two genes was based on amino acid sequence homology for the two proteins[21]. Intervening sequences are shown by the vertical bars. The size and position of the coding and intervening sequences for ovalbumin have been published elsewhere[41–43].

expected from random DNA sequences. These data provide clear evidence for the difference between the two genes in terms of position, size and composition of introns in a region of high protein sequence homology.

Evolutionary origin of the human α_1-antitrypsin and chicken ovalbumin genes

At present it is impossible to determine conclusively whether human α_1-antitrypsin and chicken ovalbumin arose by convergent or divergent evolution. The lack of similarity in structure between the two genes is certainly not unexpected if convergent evolution were the pathway. We have previously shown that overall amino acid sequence homology between a third member of this protein super-family, human antithrombin III[33], which is another plasma protease inhibitor, and α_1-antitrypsin and ovalbumin is 28% and 31% respectively[21]. The cloning and characterization of the human chromosomal antithrombin III gene will permit the comparison of the mosaic structures of these genes, which may elucidate the evolution of the gene family.

It would be very intriguing if these genes did arise by divergent evolution. In contrast to the other multi-gene families such as the globins[44] and vitellogenins[45], where the number and positions of introns are rigidly conserved, the α_1-antitrypsin and ovalbumin genes contain different numbers of introns which are located at completely different sites in the exonic sequences. This observation suggests that the ancestral gene to α_1-antitrypsin and ovalbumin may have been duplicated before the addition or deletion of intervening DNA sequences. The fact that the X-Y-ovalbumin gene family[46,47] in chicken has retained the number and positioning of the introns suggests subsequent gene duplication events involving the ancestral ovalbumin gene.

The differences in structure of the human α_1-antitrypsin and chicken ovalbumin genes resemble those in the actin gene family. The actin genes of yeast[48,49], dictyostelium[50], drosophila[51] and sea urchin[52] have one, zero, one and two introns respectively; yet although the exons have retained extensive sequence homology and resemble those of the mammalian cytoplasmic actins, the positions of the introns are completely distinct. It has thus been suggested that the current models for the functional and evolutionary conservation of locations of introns may not be valid for all genes[53], and that some introns are vestiges of transposon-like elements that have been inserted into genes, become fixed and subsequently diverged in nucleotide sequence[51]. It has also been pointed out, however, that owing to the limited numbers of introns present in the actin gene family, these genes could also have evolved by random deletion of multiple introns present in the primordial actin gene[51]. Precise deletion of an entire intron has been reported for the rat insulin gene, resulting in the existence of two genotypic alleles[54]. Similar observations have been made for the mouse α-globin gene[55]. In the case of α_1-antitrypsin and ovalbumin however, if indeed the ancestral gene to these proteins contained 10 or more introns and the two genes arose by random intron deletion, a number of introns would be expected to share common sites in the exons of the two genes.

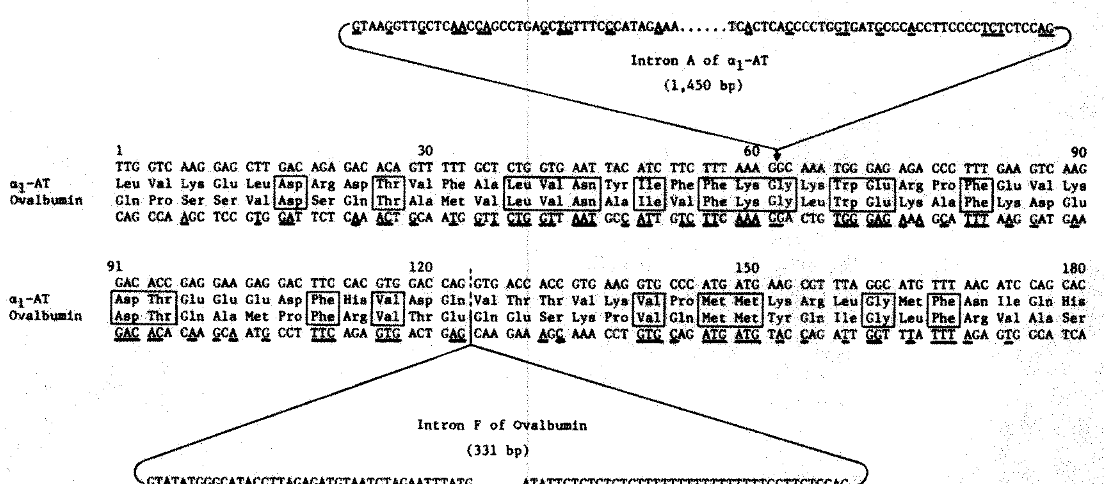


Fig. 5 Comparison of amino acid and nucleic acid sequences of splice junctions for α_1-antitrypsin (α_1-AT) intron A and ovalbumin intron F. Nucleic acid sequences corresponding to the mRNA molecules are shown in the 5'→3' direction. Identical bases for the two genes in the coding regions are underlined in the ovalbumin sequence. Forty bases at each end of each intron are shown for comparison with corresponding continuing base sequences of the other gene. Underlined bases in the introns are identical to those in the coding region of the opposite gene. The position of each intron is shown by a vertical line joining the ends of each loop. The first amino acids shown are residues 172 and 162 of mature α_1-antitrypsin and ovalbumin respectively. Ovalbumin intron E is located 81 base 5' to α_1-antitrypsin intron A (data not shown).

The fact that the positions of the introns in the two genes show no overlap makes this hypothesis less attractive. Consequently, at least some of the introns could be the vestiges of transposable elements that had been inserted into pre-existing exons of the two genes after their divergence from an ancestral gene.

We thank Dr Tom Maniatis for providing the human gene library, Rodger Staden for providing the DNA sequence computer programs and Dr Jon Herriott for adapting these programs with the local computer facility. Drs Dominic Chung and Ross MacGillivray contributed many valuable suggestions. This work was supported in part by a grant from the Cystic Fibrosis Foundation and NIH grants HL 16919 (E.W.D.) and HL 27509 (S.L.C.W). G.L.L. is the recipient of NIH Senior Research Fellowship HL 05962 and K.K. of NIH Research Career Development Award HL 00404 S.L.C.W. is an Investigator of the Howard Hughes Medical Institute.

Received 7 January; accepted 20 April 1982.

1. Heimburger, N., Haupt, H. & Schwick, H. G. in *Proc. int. Res. Conf. on Proteinase Inhibitors* (eds Fritz, H. & Tschesche, H.) 1-22 (deGruyter, Berlin, 1970).
2. Janoff, A. *Am. Rev. Resp. Dis.* **105**, 121-127 (1972).
3. Ohlsson, K. & Olsson, I. *Eur. J. Biochem.* **36**, 473-481 (1973).
4. Kueppers, F. *Am. J. hum. Genet.* **25**, 677-686 (1973).
5. Laurell, C. B. & Erikson, S. *Scand. J. clin. Lab. Invest.* **15**, 132-140 (1963).
6. Briscoe, W. A., Kueppers, F., Davis, A. L. & Bearn, A. G. *Am. Rev. resp. Dis.* **94**, 529-539 (1966).
7. Sharp, H. L., Bridges, R. A., Krivit, W. & Freier, E. F. *J. Lab. clin. Med.* **75**, 934-939 (1969).
8. Olsen, G. N., Harris, J. O., Castle, J. R., Waldeman, R. H. & Karmgard, H. J. *J. clin. Invest.* **55**, 427-430 (1975).
9. Tuttle, W. C. & Jones, R. K. *Am. J. clin. Path.* **64**, 477-482 (1975).
10. Beatty, K., Bieth, J. & Travis, J. *J. biol. Chem.* **255**, 3931-3934 (1980).
11. Bieth, J. G. *Clin. Resp. Physiol.* **16**, (Suppl.), 183-195 (1980).
12. Starkey, P. M. & Barrett, A. J. *Biochem. J.* **155**, 265-271 (1976).
13. Kuhn, C. & Senior, R. M. *Lung* **155**, 185-197 (1978).
14. Gadek, J. E. *et al. Clin. Resp. Physiol.* **16** (Suppl.), 27-40 (1980).
15. Kueppers, F. in *Lung Biology in Health and Disease* (ed. Litwin, S. D.) 23 (Dekker, New York, 1978).
16. Carrell, R. W. & Owen, M. C. *Essays med. Biochem.* **4**, 83-119 (1980).
17. Fagerhol, M. K. & Laurell, C. B. *Prog. med. Genet.* **7**, 96-111 (1970).
18. Fagerhol, M. K. *Sem. Haemat.* **1**, 153-161 (1968).
19. Allen, R. C., Harley, R. A. & Talamo, R. C. *Am. J. clin. Path.* **62**, 732-739 (1974).
20. Chandra, T., Kurachi, K., Davie, E. W. & Woo, S. L. C. *Biochem. biophys. Res. Commun.* **103**, 751-758 (1981).
21. Kurachi, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 6826-6830 (1981).
22. Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake G. & Maniatis, T. *Cell* **15**, 1156-1174 (1978).
23. Morii, M., Odani, S., Koide, T. & Ikenaka, T. *J. Biochem.* **83**, 269-277 (1978).
24. Carrell, R., Owen, M., Brennan, S. & Vaughan, L. *Biochem. biophys. Res. Commun.* **91**, 1032-1037 (1979).
25. Jackson, R. C. & Blobel, G. *Ann. N.Y. Acad. Sci.* **343**, 391-403 (1980).
26. Goldberg, M. thesis, Stanford Univ. (1979).
27. Corden, J. *et al. Science* **209**, 1406-1414 (1980).
28. Benoist, C., O'Hare, K., Breathnach, R. & Chambon, P. *Nucleic Acids Res.* **8**, 127-142 (1980).
29. Proudfoot, N. J. & Brownlee, G. G. *Nature* **263**, 211-214 (1976).
30. Hagenbuchle, O., Bovey, R. & Young, R. A. *Cell* **21**, 179-187 (1980).
31. Goeddel, D. V. *et al. Nature* **290**, 20-26 (1981).
32. Jung, A., Sippel, A. E., Greg, M. & Schutz, G. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5759-5763 (1980).
33. Hunt, L. T. & Dayhoff, M. O. *Biochem. biophys. Res. Commun.* **95**, 864-871 (1980).
34. Dayhoff M. O. (ed.) *Atlas of Protein Sequence and Structure* Vol. 5, Suppl. 3 (National Biomedical Research Foundation, Washington DC, 1979).
35. Novotny, J. *Nucleic Acids Res.* **10**, 127-131 (1982).
36. Maizel, J. V. Jr & Lenk, R. P. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
37. Gilbert, W. *Nature* **271**, 501 (1978).
38. Stein, J. P., Catterall, J. F., Kristo, P., Means, A. R. & O'Malley, B. W. *Cell* **21**, 681-687 (1980).
39. Craik, C. S., Buchman, S. R. & Beychok, S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1384-1388 (1980).
40. Sakano, H. *et al. Nature* **277**, 627-633 (1979).
41. McReynolds, L. *et al. Nature* **273**, 723-728 (1978).
42. Catterall, J. F. *et al. Nature* **257**, 510-513 (1978).
43. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. & Chambon, P. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4853-4857 (1978).
44. Maniatis, T., Fritsch, E. F., Lauer, J. & Lawn, R. M. *A. Rev. Genet.* **14**, 145-178 (1980).
45. Wahli, W., Dawid, I. B., Wyler, T., Wever, R. & Ryffel, G. U. *Cell* **20**, 107-117 (1980).
46. Colbert, D. A. *et al. Biochemistry* **19**, 5586-5592 (1980).
47. Heilig, R., Perrin, F., Gannon, F., Mandel, J. L. & Chambon, P. *Cell* **20**, 625-637 (1980).
48. Gallwitz, D. & Sures, I. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2546-2550 (1980).
49. Ng, R. & Abelson, J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3912-3916 (1980).
50. Firtel, R., Timm, R., Kimmel, A. R. & McKeown, M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 6206-6210 (1979).
51. Fyrberg, E. A., Bond, B. J., Hershey, N. D., Mixter, K. S. & Davidson, N. *Cell* **24**, 107-116, (1981).
52. Durica, D. S., Schloss, J. A. & Crain, W. R. Jr. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5683-5687 (1980).
53. Firtel, R. A. *Cell* **24**, 6-7 (1980).
54. Lomedico, P. *et al. Cell* **18**, 545-558 (1979).
55. Nishioka, Y., Leder, A. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2806-2809 (1980).
56. Benton, W. D. & Davis, R. W. *Science* **196**, 180-182 (1977).
57. Woo, S. L. C. *Meth. Enzym.* **68**, 389-395 (1979).
58. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560-564 (1977).
59. Maniatis, T., Jeffrey, A. & Kleid, D. G. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1184-1188 (1975).
60. Woo, S. L. C. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 3688-3692 (1978).
61. Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).
62. Denhardt, G. *Biochem. biophys. Res. Commun.* **23**, 641-646 (1966).
63. Smith, G. E. & Summers, M. D. *Analyt. Biochem.* **109**, 123-129 (1980).

Molecular characterization of two myosin heavy chain genes expressed in the adult heart

V. Mahdavi*, M. Periasamy* & B. Nadal-Ginard*

Albert Einstein College of Medicine, New York, New York 10461, USA

Four cDNA clones containing cardiac myosin heavy chain (MHC) inserts have been characterized. Hybridization and nucleotide sequence analysis identify two different MHC genes coding for proteins of different length that are both specifically expressed in the adult heart. The nucleotide and amino acid sequences reveal a highly conserved structure in the light meromyosin portion of MHC from striated muscle tissues.

A GROWING body of evidence suggests that several muscle-specific contractile proteins are encoded by multi-gene families. In invertebrates, birds and mammals, different actin genes[1-3] and different myosin light chains (MLCs)[4-6] are expressed in muscle and non-muscle tissues. Recent observations obtained by peptide mapping[6-8], immunological cross-reactivity[8,9] and DNA analysis[10-12] have now documented at the protein and gene levels that at least 10 myosin heavy chain (MHC) genes are expressed in the rat. In other vertebrates the MHC genes are also encoded by a multigene family[12,13]. Each muscle type (skeletal, cardiac and smooth) has tissue-specific MHCs that differ from the non-muscle (cytoplasmic) MHC[12-17]. However, the MHCs in different striated muscle tissues show a considerable degree of sequence homology not only within the same species but also across the evolutionary scale[12]. Despite these similarities, peptide and enzymatic analysis has shown that cardiac myosin is different from the myosins in other striated muscles[5,6]. The existence of distinct cardiac myosins in atrial and ventricular cells has been demonstrated in rat and chicken hearts by immunohistochemical and kinetic approaches[15,18]. Like skeletal muscle MHC[6-8], cardiac muscle MHCs seems to be developmentally regulated[6,15,19]. In several mammals, three myosin ventricular isozymes with different Ca^{2+}-ATPase activity and different MLCs and MHC composition, have been identified[6,15,19]. These three isozymes have been named V_1, V_2 and V_3 on the basis of their increasing electrophoretic mobility[15]. In the rat, a double transition between these isozymic forms is correlated with the age of the animal[15,19]. V_1, with the

* Present address: Department of Pediatrics, Harvard Medical School, Department of Cardiology, Children's Hospital Medical Center, 300 Longwood Avenue, Boston, Massachusetts 02115, USA.

0028-0836/82/250659-06$01.00

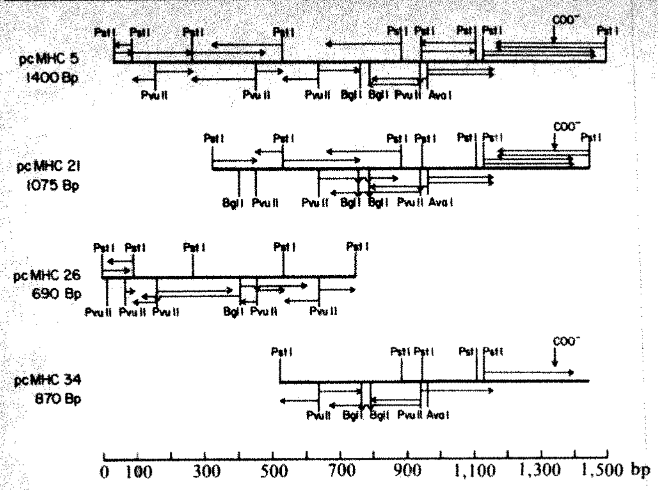


Fig. 1 Restriction endonuclease maps of cDNA clones containing MHC inserts. The recombinant cDNA clones were prepared as described elsewhere[24]. Single and double digestion with endonucleases *Pst*I, *Pvu*II, *Ava*I and *Bgl*I were in 25-μl reactions containing 6 mM Tris-HCl *p*H 7.5, 6 mM $MgCl_2$, 6 mM 2-mercaptoethanol and appropriate salt recommended by the various suppliers, for 2 h at 37°C. The relative position of the enzyme restriction sites was determined by the method of Smith and Birnstiel[42]. Solid bar represent the MHC cDNA inserts. Orientation from left to right corresponds to the 5'-3' orientation of the mRNA strand. The nucleotide length covered by the inserts is indicated. Sequencing strategy for the cMHC cDNAs clones: Horizontal arrows indicate the direction and length of each fragment sequenced. cDNA clones (20 μg) digested with the appropriate enzymes (indicated at the starting point of each arrow) were labelled at the 5' termini with [γ-^{32}P]ATP (2,000 Ci mmol^{-1}) or at the 3' termini with 3'-[α-^{32}P]dATP (cordycepin 5' triphosphate 1,000–3,000 Ci mmol^{-1}) as described elsewhere[43]. Single-end labelled fragments were obtained either by digestion with a second restriction endonuclease or by strand separation, followed by polyacrylamide gel electrophoresis. The labelled fragments were localized on the gel by X-ray autoradiography, excised from the gel and electroeluted. The DNA sequencing was performed according to the method of Maxam and Gilbert[32].

highest Ca^{2+}-ATPase activity, appears at birth and is prevalent throughout life; V_3 and the heterodimer V_2 are the only fetal forms and reappear in the second month of postnatal life. In the adult, the isozymic distribution can be modulated in different physiological and pathological conditions[20–23]. The differences seen in the polypeptide composition of the MLCs and MHCs of embryonic and adult hearts suggest that different myosin genes are expressed at specific developmental stages of the rat[6,21].

To improve our understanding of the molecular mechanisms responsible for the developmental and tissue-specific expression of the cardiac MHC genes and the possible physiological and pathological implications of the isozymic transitions, we have constructed recombinant cDNA clones that contain cardiac MHC mRNA sequences. We show here that at least two different but closely related cardiac MHC genes are expressed in the adult rat. These two genes are apparently developmentally regulated. The structure and nucleotide sequence of these cardiac MHC cDNA clones are discussed and compared with those of skeletal muscle MHCs.

Isolation of MHC cDNA clones from cardiac muscle

The extent of cross-homology within the MHC gene family[12] makes it difficult to identify the MHC gene(s) expressed in different cell types and at a particular time of development. Therefore, identification of gene-specific sequences becomes essential when studying MHC gene regulation at the molecular level. To this end, we constructed a cDNA library by the G – C tailing method in *Pst*I site of pBR322 using standard cloning techniques[24], with poly(A)$^+$ RNA isolated from the ventricles of a 3-month-old rat (Wistar strain) as template for cDNA synthesis. We identified 16 recombinant plasmids containing MHC inserts, by the Grunstein and Hogness colony hybridization procedure[25], using as probe for MHC sequences the 320-base pair (bp) internal *Pst*I restriction fragment of the embryonic skeletal muscle MHC cDNA clone $pMHC_{25}$ (ref. 24).

It has been previously shown that $pMHC_{25}$, obtained from L_6E_9 myotubes[26,27], contains sequences that are highly conserved within the sarcomeric MHC gene family and cross-hybridizes to cardiac mRNA[10,12]. However, this procedure limited the detection to those cardiac MHC cDNA clones that cross-hybridize to the embryonic skeletal MHC and contain the region corresponding to the insert used as probe. Therefore, the 16 cDNA clones represent a minimum estimate of the MHC cDNA clones in the cDNA library.

The four cardiac MHC cDNA clones that contain the longest inserts (690–1400 nucleotides) were selected for further analysis. Note that because MHC mRNA is 7,100 nucleotides long[24], the longest clone, $pCMHC_5$, contains 23% of the MHC mRNA sequence. As shown in Fig. 1, $pCMHC_5$, $pCMHC_{21}$, $pCMHC_{26}$ and $pCMHC_{34}$ have very similar restriction endonuclease maps and have been aligned to facilitate comparison. A remarkable feature of these maps is the frequency of sites for the restriction endonucleases *Pst*I and *Pvu*II that recognize the nucleotide sequences CTGCAG (Leu–Gln) and CAGCTG (Gln–Leu), respectively. This feature seems to be a characteristic of the vertebrate MHC genes so far studied[13,24,28,29]. The *Pst*I and *Ava*I sites are conserved in the region where the different clones overlap. Taking $pCMHC_5$ as a reference, $pCMHC_{21}$ has an additional *Bgl*I site at coordinate 400 of the diagram. $pCMHC_{26}$ also has a *Bgl*I site at position 400 and an additional *Pvu*II site at position 80. $pCMHC_{34}$ covers a region of 900 nucleotides that shows identical restriction enzyme sites in all four cDNA clones. On the basis of these observations, the MHC cDNA clones represented in the cardiac cDNA library cover approximately the same portion of MHC mRNA and can be divided into two classes, one represented by $pCMHC_{21}$ and $pCMHC_{26}$ and the other by $pCMHC_5$; $pCMHC_{34}$ could belong to either class. Each of these two sets probably originates from a single mRNA species, thus suggesting a minimum of two cardiac MHC genes expressed in the ventricular myocardium of the heart used for the cDNA cloning.

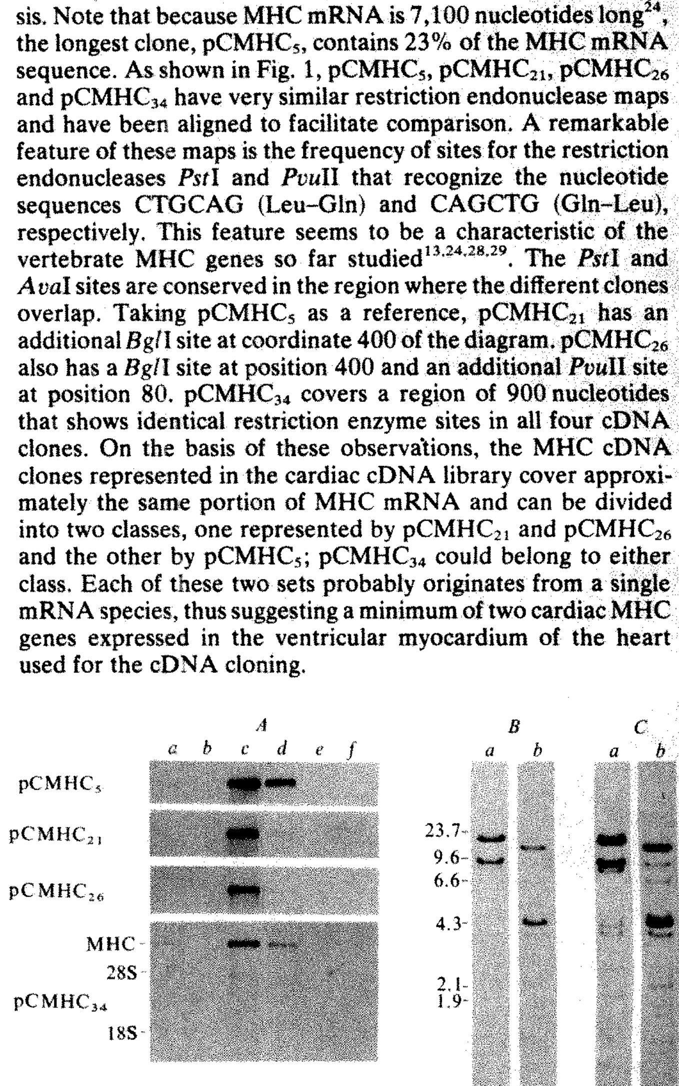


Fig. 2 Identification of MHC cDNA sequences in RNA from different rat tissues and in rat genomic DNA. Total cytoplasmic RNAs were isolated from different tissues by the hot phenol procedure[44] and from L_6E_9 cells by using guanidine hydrochloride[45,46]. RNA (5 μg per lane) was fractionated on 35% formaldehyde, 1% agarose gels and blotted onto nitrocellulose[47]. DNA was isolated from rat liver nuclei[48], and 10 μg were digested with restriction endonucleases in 50 μl reaction buffer (see Fig. 1 legend), separated on 1% agarose-TAE buffer (40 mM Tris-acetate *p*H 7.8, 5 mM Na-acetate, 1 mM EDTA) and transferred onto nitrocellulose[30]. cDNA probes were radiolabelled by nick-translation in the presence of [α ^{32}P]CTP and [α^{32}P]TTP (400 Ci mmol^{-1})[31]. 10^6 c.p.m. were used per gel lane for hybridization. Hybridization washes were in 0.1 × SSC, 0, 1% SDS at 65 °C. *A*, hybridization of $pCMHC_5$, $pCMHC_{21}$, $pCMHC_{26}$, $pCMHC_{34}$ to rat RNA from: *a*, L_6E_9 myotubes; *b*, L_6E_9 myoblasts; *c*, adult cardiac muscle; *d*, adult skeletal muscle; *e*, uterine smooth muscle; *f*, NRK fibroblasts. Marker indicates migration of 28S and 18S ribosomal RNAs. *B*, hybridization of $pCMHC_5$ to Wistar rat DNA digested with: *a*, *Eco*RI; *b*, *Hin*dIII. Size markers indicated in kilobases are from *Hin*dIII-digested λ phage DNA. Exposure of the autoradiogram was for 3 days. *C*, same as in *B* except that exposure of the autoradiogram was for 14 days.

```
                                  27                                 54
Asp Leu Thr Glu Gln Leu Gly Glu Gly Gly Lys Asn Val His Glu Leu Glu Lys
GAC CTG ACT GAA CAG CTG GGA GAA GGC GGT AAA AAT GTG CAC GAG CTG GAG AAG
                                                    -------------------
                                                    -------------------

                                  81                                108
Ile Arg Lys Gln Leu Glu Val Glu Lys Leu Glu Leu Gln Ser Ala Leu Glu Glu
ATC CGC AAA CAG CTG GAG GTG GAG AAG CTG GAA CTG CAG TCA GCC CTA GAA GAG
G-----A--GG---A----------CT---------------G----------------------------
Val-----Arg-------------Ala--------------------------------------------

                                  135                               162
Ala Glu Ala Ser Leu Glu His Glu Glu Gly Lys Ile Leu Arg Ala Gln Leu Glu
GCT GAG GCC TCC CTG GAG CAT GAG GAG GGC AAG ATC CTC CGA GCC CAG CTG GAG
-----------------------------------------------------------------------
-----------------------------------------------------------------------

                                  189                               216
Phe Asn Gln Ile Lys Ala Glu Ile Glu Ser Lys Leu Ala Glu Lys Asp Glu Glu
TTC AAC CAG ATC AAG GCA GAG ATC GAA AGC AAG CTG GCA GAG AAG GAC GAG GAG
--------------------------------------G--------------------------------
------------------------------------Arg--------------------------------

                                  243                               270
MET Glu Gln Ala Lys Arg Asn His Leu Arg Val Val Asp Ser Leu Gln Thr Ser
ATG GAG CAG GCC AAG CGC AAC CAC CTG CGG GTG GTG GAC TCC CTG CAG ACC TCC
-----------------------------------------------------------------------
-----------------------------------------------------------------------

                                  297                               324
Leu Asp Ala Glu Thr Arg Ser Arg Asn Glu Ala Leu Arg Val Lys Lys Lys MET
CTA GAT GCC GAG ACG CGC AGC CGC AAC GAG GCC CTG CGG GTG AAG AAG AAG ATG
--G--------------------------------------------------------------------
-----------------------------------------------------------------------

                                  351                               378
Glu Gly Asp Leu Asn Glu MET Glu Ile Gln Leu Ser Gln Ala Asn Arg Ile Ala
GAG GGC GAC CTC AAT GAG ATG GAG ATC CAG CTC AGT CAG GCC AAT AGA ATA GCC
------------------C-------------------------------T-------C-C-C---G---T
------------------------------------------------His-------------Met----

                                  405                               432
Ser Glu Ala Gln Lys His Leu Lys Asn Ala Gln Ala His Leu Lys Asp Thr Gln
TCA GAG GCC CAG AAG CAC TTG AAG AAC GCC CAA GCC CAC TTG AAG GAC ACC CAG
G-T---------------A---A-G--------G--CT----G-AGT-TTG-C-------------T---A
Ala-----------------Gln-Val-----Ser-Leu-----Ser-Leu--------------------

                                  459                               486
Leu Gln Leu Asp Asp Ala Val Arg Ala Asn Asp Asp Leu Lys Glu Asn Ile Ala
CTC CAG CTG GAT GAC GCA GTC CGT GCC AAT GAC GAC CTG AAG GAG AAC ATC GCC
A----------------------------------------------------------------------
Ile--------------------------------------------------------------------

                                  513                               540
Ile Val Glu Arg Arg Asn Thr Leu Leu Gln Ala Glu Leu Glu Glu Leu Arg Ala
ATC GTG GAG CGG CGC AAC ACC CTG CTG CAG GCG GAG CTG GAG GAG CTG CGG GCC
-------------------------A---------------------------------------------
------------------------Asn--------------------------------------------

                                  567                               594
Val Val Glu Gln Thr Glu Arg Ser Arg Lys Leu Ala Glu Gln Glu Leu Ile Glu
GTG GTG GAG CAG ACA GAG CGG TCT CGG AAG CTG GCA GAG CAG GAG CTG ATC GAG
------------------G----------------------------------------------------
-----------------------------------------------------------------------

                                  621                               648
Thr Ser Glu Arg Val Gln Leu Leu His Ser Gln Asn Asn Ser Leu Ile Asn Gln
ACC AGC GAG CGC GTG CAG CTG CTG CAC TCC CAG AAC AAC AGC CTC ATC AAC CAG
--------------G----------------------------------C---------------------
------------------------------------------------Thr--------------------

                                  675                               702
Lys Lys Lys MET Asp Ala Asp Leu Ser Gln Leu Gln Thr Glu Val Glu Glu Ala
AAG AAG AAG ATG GAT GCA GAC CTC TCC CAG CTC CAG ACA GAG GTG GAG GAG GCG
----------------C------------------------------------------------------
----------------His----------------------------------------------------

                                  729                               756
Val Gln Glu Cys Arg Asn Ala Glu Glu Lys Ala Lys Lys Ala Ile Thr Asp Ala
GTG CAG GAG TGT AGG AAC GCA GAG GAG AAG GCC AAG AAG GCC ATC ACA GAT GCC
-----------------------------------------------------------------------
-----------------------------------------------------------------------

                                  783                               810
Ala MET MET Ala Glu Glu Leu Lys Lys Glu Gln Asp Thr Ser Ala His Leu Glu
GCC ATG ATG GCC GAG GAG CTG AAG AAG GAG CAG GAC ACC AGC GCC CAC CTA GAG
------------------------------------------------------------------G----
-----------------------------------------------------------------------

                                  837                               864
Arg MET Lys Lys Asn MET Glu Gln Thr Ile Lys Asp Leu Gln His Arg Leu Asp
CGC ATG AAG AAG AAC ATG GAG CAG ACC ATC AAG GAC CTG CAG CAC CGC CTG GAC
--------------------------------------------------------------G--------
-----------------------------------------------------------------------

                                  891                               918
Glu Ala Glu Gln Ile Ala Leu Lys Gly Gly Lys Lys Gln Leu Gln Lys Leu Glu
GAG GCA GAG CAG ATC GCC CTC AAG GGT GGC AAG AAG CAG CTG CAG AAG CTG GAG
-----------------------------------------------------------------------
-----------------------------------------------------------------------

                                  945                               972
Ala Arg Val Arg Glu Leu Glu Asn Glu Leu Glu Ala Glu Gln Lys Arg Asn Ala
GCC CGG GTC CGC GAG CTG GAG AAT GAG CTG GAG GCT GAG CAG AAG CGC AAT GCG
-----------------------------------------------------------------------
-----------------------------------------------------------------------

                                  999                              1026
Glu Ser Val Lys Gly MET Arg Lys Ser Glu Arg Arg Ile Lys Glu Leu Asn Tyr
GAG TCG GTG AAG GGC ATG AGG AAG AGC GAG CGG CGC ATC AAG GAG CTC AAC TAC
-----------------------------------------------------------------------
-----------------------------------------------------------------------

                                  1053                             1080
Gln Thr Glu Glu Asp Lys Lys Asn Leu Val Arg Leu Gln Asp Leu Val Asn Lys
CAG ACA GAG GAA GAC AAG AAG AAC TTG GTG CGG CTG CAG GAC CTG GTG AAC AAG
--------------------------------C-A-C-----A---------------------G------
------------------------------------Leu-------------------------Asp----

                                  1107                             1134
Leu Gln Leu Lys Val Lys Ala Tyr Lys Arg Gln Ala Glu Glu Ala Glu Glu Gln
CTG CAG TTG AAG GTG AAG GCC TAC AAG CGC CAG GCT GAA GAG GCG GAG GAA CAG
-----------------------------------------------------------------------
-----------------------------------------------------------------------

                                  1161                             1188
Ala Asn Thr Asn Leu Ser Lys Phe Arg Lys Val Gln His Glu Leu Asp Glu Ala
GCC AAC ACC AAC CTG TCC AAG TTC CGC AAG GTG CAG CAC GAG CTG GAT GAG GCA
-----------------------------------------------------------------------
-----------------------------------------------------------------------

                                  1215                             1242
Glu Glu Arg Ala Asp Ile Ala Glu Ser Gln Val Asn Lys Leu Arg Ala Lys Ser
GAG GAG AGG GCG GAC ATT GCC GAG TCC CAG GTC AAC AAG CTG CGG GCC AAG AGC
-----------------------------------------------------------------------
-----------------------------------------------------------------------

                                  1269                             1297
Arg Asp Ile Gly Ala Lys Gln Lys Ile His Asp Glu Glu END
CGT GAC ATT GGC GCC AAG CAG AAA ATC CAC GAT GAG GAA TAA CCTGTCCAGCAGAAA
------------------------GGC-CTG--AT-G-A---G-T---ATC--TG--TCTA--CAACCCT-
------------------------Gly Leu Asn Glu Glu END

                                  1334                             1370
GAGCCTCGCCGTTGCCATCCCACAATAAATACGAATGTTCG (A)27
AG-ATG-CTGTGAAG-CCTCAGACC-GG-GC-TTTGAAA-AGCACCTTCAGGCAGAAACACAATAAAGCAA

TTTTCCTTCAAGCC (A)37
```

Fig. 3 Nucleotide sequence of cardiac MHC cDNA clone inserts. The Maxam and Gilbert[32] method was used for sequence determination. The sequencing strategy is depicted in Fig. 1. $pCMHC_{26}$ (nucleotides 1–790) $pCMHC_{21}$ (nucleotides 350–1,340) and $pCMHC_{34}$ (nucleotides 500–1,340) sequences that represent portions of the same cardiac MHC gene product have been joined into a continuous 1,340-nucleotide sequence. In $pCMHC_{34}$, the longer nucleotide sequence of the untranslated 3′ end (nucleotides 1,341–1,351) reads ATTTTGCCTGC(A)90. The nucleotide and the derived amino acid sequences are written 5′ to 3′ and NH_2 to COOH, respectively, in the upper two lines. The nucleotide and amino acid sequences of $pCMHC_5$ that represent the other cardiac MHC gene product (nucleotides 40–1,384) are in the bottom two lines. In $pCMHC_5$, nucleotides and amino acid are written only where different from that of $pCMHC_{26,21}$ and $_{34}$. The dashed line indicates identical sequence to that of $pCMHC_{21,26,34}$.

The MHC cDNA clones are specific for cardiac MHC sequences

Because each muscle and non-muscle tissue has distinct MHCs[12–17], we investigated the tissue specificity of the four MHC cDNA clones. These cDNA clones were hybridized individually to electrophoretically separated RNA extracted from different striated muscle types (cardiac and skeletal), smooth muscle (uterine) and non-muscle (fibroblast) tissues. Figure 2*A* shows that all four cDNA clones hybridize to a 31S MHC mRNA from striated muscle tissues (lanes *a*, *c*, *d*) and do not hybridize to smooth muscle nor to non-muscle tissue RNA (lanes *b*, *e*, *f*). The same pattern of hybridization is observed when poly(A)$^+$ RNA is used instead of total RNA (data not shown). Therefore, the lack of hybridization of the cardiac MHC cDNA clones to non-sarcomeric RNA is probably due to absence of nucleotide sequence homology between these MHC mRNA sequences rather than differences in the amounts of MHC mRNA in the RNA samples[17]. Hybridization is always strongest with cardiac MHC mRNA from the adult rat (lane *c*). Hybridization to L_6E_9 myotube and skeletal MHC mRNA (lanes *a*, *d*) could reflect cross-homology of the cardiac MHC cDNA clones to common sequences in these different MHC mRNAs (see below and ref. 12). These results indicate that all four MHC cDNA clones are cardiac specific but do not exclude the possibility that the MHC mRNAs corresponding to these clones are present in L_6E_9 myotubes and skeletal muscle in low amounts.

Genomic distribution of the cardiac MHC sequences

A previous study on the genomic organization of the MHC sequences has determined that a minimum of eight genes encoding striated muscle MHCs are detected in the rat genome[10]. We attempted to determine which and how many striated MHC genes could be identified as cardiac specific. The four cardiac cDNA clones were ^{32}P-labelled[30] and hybridized individually to restriction digests of rat DNA (Wistar strain) prepared by the Southern blotting technique[31]. $pCMHC_5$ (Fig. 2, panel *B*) and $pCMHC_{21,26}$ and $_{34}$ (not shown) generate the same pattern when hybridized to genomic DNA. At high stringency, this pattern consists mainly of two strongly hybridizing fragments

of approximately 8 and 21 kilobases (kb) in the *Eco*RI digest (lane *a*) and 4.5 and 18 kb in the *Hin*dIII digest (lane *b*). The other hybridization bands that become more apparent on long exposure of the autoradiograms (Fig. 2, panel *C*) reflect cross-hybridization to segments corresponding to other skeletal muscle genes[10]. However, due to the poor resolution of the gel system within this size range it is not possible to ascertain whether each band represents only one or more gene fragments.

Considering that the cardiac MHC cDNA clones have neither *Eco*RI nor *Hin*dIII restriction sites, it is probable that each pair of strongly hybridizing fragments represents indeed two different cardiac MHC genes. Alternatively, the hybridization pattern could be produced by a single MHC gene, interrupted by an intervening sequence containing the appropriate restriction sites. This possibility is very unlikely because internal *Pst*I restriction fragments of $pCMHC_5$ and $pCMHC_{21}$ generate the same hybridization pattern as the entire cDNA clones (data not shown).

This result suggests that more than one cardiac MHC gene is represented in the genome. However, a clear determination of how many and which cardiac genes encode for the MHC mRNAs represented by cDNA clones is hampered by the obvious sequence homology within these genes.

Nucleotide sequence analysis of the cardiac MHC cDNA clones

To obtain unequivocal information on the number of cardiac MHC genes represented in the MHC cDNA clones, understand the extent of their homologies and identify gene-specific sequences, we determined the complete nucleotide sequence of the four cDNA clones by the method of Maxam and Gilbert[32].

The sequence analysis shows that the cDNA clones comprise the sequence of the 430 carboxyl terminal amino acids, the complete untranslated 3′ end and a portion of the poly(A) track of MHC mRNA. These four clones represent the transcription products of two distinct MHC genes that have 95% homology in the coding region but are clearly different in their untranslated 3′ ends. The sequences of $pCMHC_{21}$ $pCMHC_{26}$ and $pCMHC_{34}$ overlap over 400 base pairs and clearly represent transcription products of the same MHC gene. However, in $pCMHC_{34}$, the untranslated 3′ end is 11 nucleotides longer than that of $pCMHC_{21}$. This extra sequence precedes the poly(A) track of the MHC mRNA (see Fig. 3 legend), thus indicating heterogeneity in the polyadenylation site. The sequence of $pCMHC_5$ corresponds to a segment of a different MHC gene product.

The continuous nucleotide and derived amino acid sequences of these two MHCs, one represented by $pCMHC_{21}$, $pCMHC_{26}$ and $pCMHC_{34}$ and the other by $pCMHC_5$ are shown in Fig. 3. A characteristic of this portion of the two cardiac MHC mRNAs

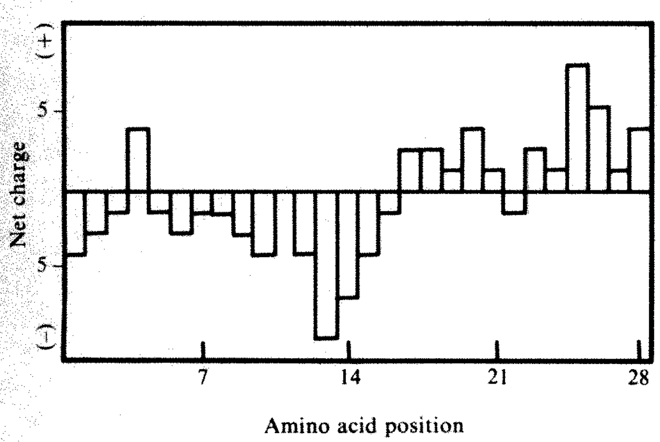


Fig. 4 Distribution of charged residues in the LMM portion of the cardiac MHC 'encoded' by $pCMHC_{21/26}$. The LMM was divided into periods of 28 residues. The relative position of each amino acid within the period is represented on the *x* axis. Positive and negative net charge at each position is recorded on the *y* axis.

is the high G+C content (33.7% G and 26.9% C) and the low U content (11.2%). This feature, together with the presence of several inverted repeats, allows the formation of highly stable hairpin structures ($\Delta G = -12$ to -17.5 kilocalories). A computer search did not reveal the presence of significant direct repeats in either sequence.

Comparison of the two cardiac MHC mRNA sequences shows very close similarities in this portion of the molecules. Long stretches of almost complete identity are interrupted by single nucleotide changes that occur at 25 to 100 base intervals. A long segment of conserved sequence, with only two base substitutions, extend from nucleotides 663 to 1,050. However, the two sequences are strikingly different in two regions: from nucleotide 360 to 433, where there is a 30% base mismatch, and from nucleotide 1,260 to the end of the molecules where the sequences completely diverge. This 3′ end divergence has several interesting features. In $pCMHC_{21/26}$ this segment, which codes for a longer MHC protein with two extra amino acids, includes the last 21 nucleotides of the coding sequence, the termination codon UAA and a short untranslated 3′ end of 60 nucleotides. Moreover, in this mRNA, the poly(A) track starts 10 nucleotides after the AATAAA sequence, thought to be the poly(A) addition signal[33]. In $pCMHC_5$, the divergent 3′ coding portion is only 15 nucleotides long. Furthermore, the termination codon UAG precedes an untranslated sequence of 100 nucleotides. In this mRNA the poly(A) track starts 18 nucleotides after the AATAAA sequence. The drastic differences in nucleotide sequence and in length of the 3′ ends of the two molecules confirm unambiguously that at least two cardiac MHC genes are expressed in the adult rat.

Amino acid sequence derived from the cardiac MHC cDNA clones

The coding sequence of $pCMHC_{21/26}$ and $pCMHC_5$ represent the 430 carboxyl-terminal amino acids of two MHCs. This portion of the molecule forms an α-superhelical coiled-coil dimeric structure[34] known as the light meromyosin (LMM). The LMM amino acid sequence of these two cardiac MHCs have 97% homology. Moreover, there are striking similarities between these sequences and rat[29] and rabbit skeletal MHCs[34]. The amino acid composition derived from the two cardiac sequences is characteristic of LMM[34]. Half of the sequence is constituted by five amino acids (Glu, Gln, Leu, Lys, Ala). As expected, there are no Pro residues that disrupt the α-helix. The codon utilization frequency is strongly biased for G or C in the third position. For instance, Glu is coded 89% of the time by GAG; Gln, 98% CAG; Leu, 65% CTG; Lys, 90% AAG and ALA, 62% GCC. It merits mention that the codon usage for the body wall MHC in *Caenorhabditis elegans* is different from that of the rat cardiac MHCs, even though these two proteins show substantial homology (J. Karn, personal communication).

Fifty per cent of the nucleotide substitutions between the two cardiac MHCs occur at the third position and do not change the corresponding amino acid (Fig. 3). The few amino acid changes observed are conservative in hydrophobicity and/or charge. This is so even in the two portions of the molecule where most amino acid substitutions are clustered: residues 120–150 in $pCMHC_{21/26}$ and the carboxyl-terminal ends of the two cardiac MHCs. Although the carboxyl-terminal sequences differ in length and diverge in amino acid composition, both terminate with two glutamic acid residues (Gln-Lys-Ile-His-Asp-Glu-Glu in $pCMHC_{21}$ and Gly-Leu-Asn-Glu-Glu in $pCMHC_5$). The carboxyl-terminal Glu-Glu is a common feature for all MHC sequences so far obtained[29,34].

From a structural point of view, the cardiac LMM sequences present a characteristic periodicity in the distribution of non-polar groups and charged residues. Non-polar, hydrophobic residues follow in general a 3,4,3,4 spacing, conforming to a stabilized α-helical structure as described for the tropomyosin molecule[35] and other MHCs[34]. Occasional deviation of this

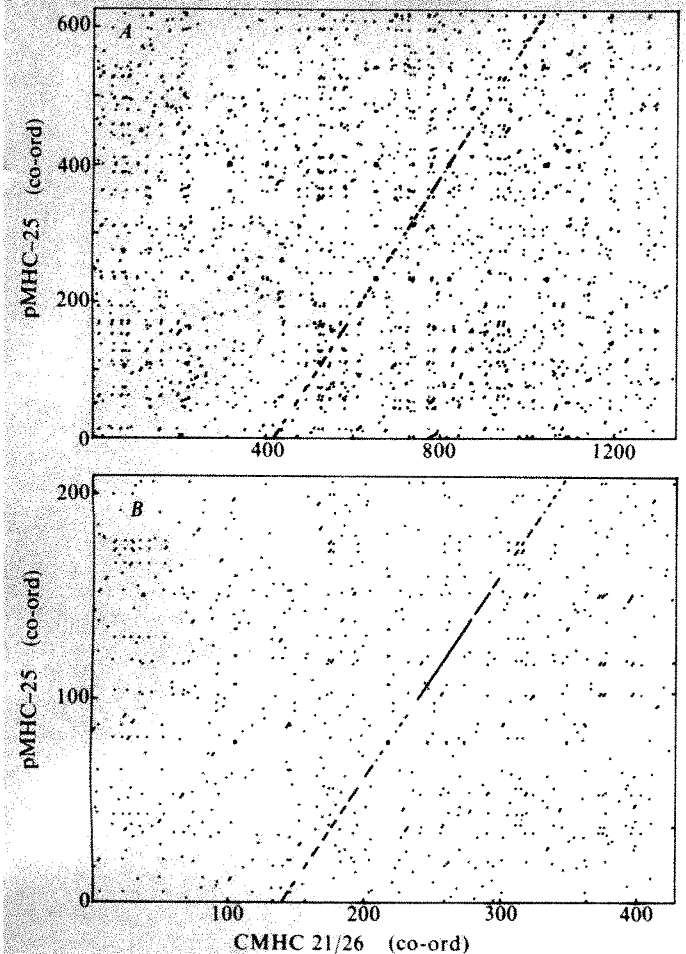


Fig. 5 Computer graphic representation of sequence homology in $pCMHC_{21/26}$ and $pMHC_{25}$. A computer program was designed to detect and represent graphically homology between two sequences displayed on the x and y axis, respectively (R. Medford, unpublished). The computer compares sequentially a given number of nucleotides (or amino acids) from the sequence displayed on the x axis to each position in the sequence on the y axis. A homology match will generate a dot at the corresponding coordinate. Continuous homology is represented by a diagonal segment covering the length of the entire corresponding sequence. Displaced sequence identities, outside the main homology axis, are represented by short diagonal segments at their relative positions. *A*, comparison of the nucleotide sequence of $pCMHC_{21/26}$ and $pMHC_{25}$. $pMHC_{25}$ covers nucleotides 415–1,045 of the sequence of $pCMHC_{21/26}$. Homology of five nucleotides in a row is required for a dot placement. This setting permits 50% third-base substitutions to be considered as homology match. *B*, comparison of the amino acid sequence of $pCMHC_{21/26}$ and $pMHC_{25}$. An homology match is two amino acids in a row.

pattern, where the spacing is 3,3,4,4, is also observed. Clusters of positively and negatively charged residues are well separated and consist in most cases of sequential repeats of the same amino acid, for instance, Lys-Lys-Lys (positive charge) and Glu-Glu-X-Glu (negative charge). Charged clusters of identical sign show an optimal periodicity of 28 residues (see Fig. 4).

These structural features, which are conserved in rat[28,29] and rabbit[34] skeletal muscle MHCs, as well as in body wall MHC of *C. elegans* (J. Karn and A. D. McLachlan, personal communication), are probably relevant to the formation and maintenance of the coiled-coil secondary structure of the LMM and to the organization of the myosin molecules in the thick filament. Interaction between charged groups within adjacent myosin chains might confer stabilizing properties for the assembly of the thick filament. Computer modelling experiments on the MHC in *C. elegans* have shown (J. Karn and A. D. McLachlan, personal communication) that surface charges, clustered on the outer surface of the α-helix, are best neutralized when two myosin filaments, aligned in parallel, are shifted by 98 residues. This would produce a 146 Å axial displacement of the myosin molecules. The displacement agrees very closely with the X-ray diffraction data showing myosin cross-bridges in the thick filament, arranged in a helical configuration with a 143 Å periodicity[36].

Cardiac and skeletal muscle LMMs have highly homologous sequences

Sequence conservation in MHC genes from striated muscle is recognizable not only within the same species but also among species as distant as nematode, sea urchin, rat and human[12]. To explore the distribution and the extent of sequence conservation between two MHCs from different striated muscles, we compared the sequences of cardiac MHC cDNA ($pCMHC_{21/26}$) and embryonic skeletal MHC cDNA ($pMHC_{25}$)[28], using a computer program designed to analyse and represent graphically sequence homologies (R. Medford, unpublished). The computer graphic display of such analysis is shown in Fig. 5*A*. The two sequences show close similarities as represented by the homology axis. Highly conserved domains, illustrated by continuous segments, are interspersed with regions of non-homology, represented by interruptions in the diagonal line.

One half of the restriction fragment of $pMHC_{25}$, which was used as a probe to screen the cardiac cDNA clones and other MHC cDNA and genomic clones (nucleotides 300–450 in $pMHC_{25}$), is 92% homologous to the cardiac clones and to other MHC cDNA clones from skeletal muscle[29].

The computer analysis of the amino acid sequence comparison of the two proteins 'encoded' by $pMHC_{25}$ and $pCMHC_{21/26}$ is presented in Fig. 5*B*. Comparison of Fig. 5*A* and *B* clearly shows that amino acid sequence homology between these two cDNA clones is higher than the nucleotide sequence homology (compare nucleotides 300–450 with amino acids 100–150 on the $pMHC_{25}$ axis). Moreover, distribution of charged and non-polar residues is conserved in the segment of protein encoded by $pMHC_{25}$ (ref. 28). Structural and functional requirements of the sarcomeric MHC might account for such a feature.

The cardiac MHC genes are developmentally regulated

Age-dependent changes of cardiac myosin isozymes have been observed in several mammals[15,16,18]. In the 12-week-old rat, cardiac myosin is 80% V_1 while the fetal forms V_3, and the heterodimer V_2, constitute the remaining 20% of the myosin[18]. As the cardiac MHC cDNA clones were constructed from cardiac mRNA of a 3-month-old animal, one set of cDNA

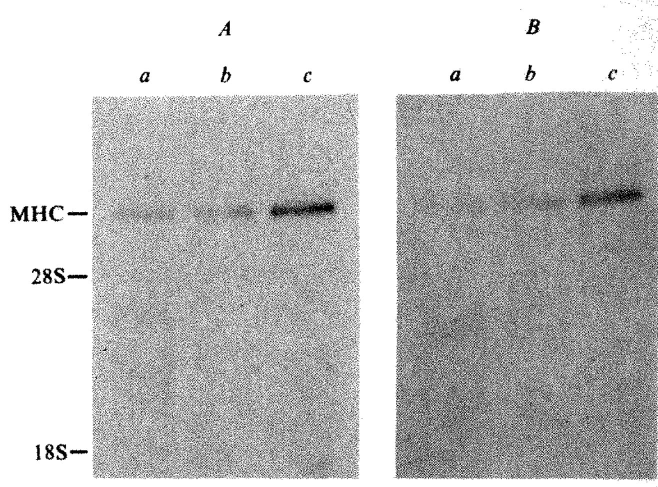


Fig. 6 Identification of cardiac MHC cDNA clones in rat cardiac RNA from different stages of development. Hybridization of $pCMHC_5$ (*A*) and $pCMHC_{21}$ (*B*) to rat cardiac RNA from: *a*, 20-day-old fetuses; *b*, 1-day newborn; *c*, 3-month adult. Samples were prepared as described in Fig. 2 legend. Hybridization washes were in 0.1×SSC, 0.1% SDS at 55 °C.

clones could represent the MHC in the major adult cardiac myosin V_1 and the other set could represent the MHC in the fetal cardiac myosin V_3. To investigate this possibility, the cDNA clones were hybridized to equal amounts of cardiac RNA isolated from rats at different stages of development, when either V_1 or V_3 myosin isozyme prevails[15,18]. The result of such an experiment (Fig. 6) shows that both $pCMHC_5$ and $pCMHC_{21}$ (each containing the 3′ end of the respective mRNA sequence) hybridize preferentially to MHC mRNA from the adult heart (panels *A* and *B*, lane *c*) and hybridize weakly to fetal and newborn RNA (panels *A* and *B*, lanes *a* and *b*). In fact, both entire clones, as well as their respective untranslated 3′ ends (data not shown), show an increase in intensity of hybridization from the 20-day-old fetus and 1-day-old animal (lanes *a*, *b*) to adult MHC mRNA (lane *c*), which closely agrees with the reported data on the ontogeny of the V_1 isozyme[15,18]. However, neither $pCMHC_5$ nor $pCMHC_{21}$ hybridizes preferentially to the 20-day-old fetal cardiac mRNA where 70% of the myosin is presumably the embryonic type V_3. Based on the assumption that the steady-state level of MHC mRNA, as a relative fraction of the total RNA population, does not change dramatically during heart development, these observations strongly suggest that in the rat not one, but at least two different adult-specific MHC genes are expressed in the ventricular myocardium. The weak hybridization to fetal MHC mRNA could reflect expression of these two cardiac MHCs early in development.

This conclusion is not necessarily in contradiction with previous reports on the existence of only one adult specific cardiac MHC isozyme[15,18]. Based on the amino acid sequence, the two adult cardiac MHCs described here would be undistinguishable, at least in the LMM portion, by any of the techniques used so far to characterize the cardiac MHC at the protein level. Whether the high degree of sequence homology between these two MHC extends throughout the entire length of the molecule remains to be determined.

Discussion

The results reported here clearly show that a minimum of two adult-specific cardiac MHC genes are expressed in the ventricles. The existence of these two MHC genes raises several questions. What is their biological and physiological significance? Are these two genes expressed, in a coordinated fashion, in the same cell? At what precise stage of the heart development and at what level are they being transcribed? Are they expressed in other muscle tissues? Some of these questions can now be approached by using the gene-specific sequences at the untranslated 3′ termini of the mRNAs.

Class switches from fetal to adult gene(s) have been described for globin[37], skeletal muscle MHC[8] and other systems[38]. The isozymic transition of cardiac MHC during development requires modulation of the expression of at least two sets of MHC genes (fetal and adult). The remarkable feature of the cardiac MHC genes is that, in the rat, as well as in several other mammals[18], the level of expression of fetal and adult MHC genes varies throughout the lifespan of the animal. Moreover, the expression of these genes can also be modulated in a reversible fashion by physiological conditions such as mechanical overload and level of circulating hormones[15,21–23,39–41]. The adaptative regulation of expression of the MHC genes in cardiac tissue, as well as the existence of several MHC genes in other striated muscles[8,10–13] offers an excellent model in which to study the control mechanisms that determine quantitative, temporal and spacial expression of specific members of the MHC gene family.

There is an apparent paradox in the multiplicity of the MHC genes and their intra- and interspecies sequence homology. The nucleotide sequence of the adult cardiac, embryonic and adult skeletal MHC mRNAs[28,29] suggests that the sarcomeric MHC genes have evolved by duplication of a common ancestor gene. Furthermore, the amino acid sequence homology and the conservation of charged and non-polar residues indicate that secondary and possibly tertiary structure is very similar in the LMM portion of the different rat[28,29] and rabbit[34] MHCs. These features indicate that the MHC genes are under evolutionary pressure for the conservation of certain sequences that might be relevant to the structure of the sarcomere. However, the large number of MHC genes points toward precise structural and/or kinetic requirements of each muscle, at various stages of development, that are best fulfilled by the corresponding isozyme. To understand the biological meaning of the MHC multi-gene family and its significance to muscle physiology, the structural and biochemical properties, as well as the subcellular location of each MHC isozyme, need to be determined.

We thank Dr J. Karn for sharing his results before publication and for helpful discussions, Dr D. Wieczorek for the Southern blot in Fig. 2*B*,*C*, Dr L. A. Leinwand who helped us to set up the DNA sequencing procedure, and Dr R. M. Medford, J. Metlay and Dr R. M. Wydro for the computer analysis. V.M. was supported by a New York Heart Association research award. This work was funded in part by grants from the NIH and the NYHA.

Received 1 November 1981; accepted 8 April 1982.

1. Reviewed in Firtel, R. A. *Cell* **24,** 6–7 (1981).
2. Engle, J. N., Gunning, P. W. & Kedes, L. *Proc. natn. Acad. Sci. U.S.A.* **78,** 4674–4678 (1981).
3. Shani, M., Zevin-Sonkin, Saxel, O., Carmon, Y., Katcoff, D., Nudel, U. & Yaffe, D. *Devl Biol.* **86,** 483–492 (1981).
4. Lowey, S., Benfield, P. A. Silberstein, L. & Lang, L. M. *Nature* **282,** 522–524 (1979).
5. Hoh, J. F. Y. & Yeoh, G. P. S. *Nature* **280,** 321–323 (1980).
6. Whalen, R. G. & Sell, S. M. *Nature* **286,** 731–733 (1980).
7. Whalen, R. G., Schwartz, K., Bouveret, P., Sell, S. M. & Gros, F. *Proc. natn. Acad. Sci. U.S.A.* **76,** 5197–5201 (1979).
8. Whalen, R. G. *et al. Nature* **292,** 805–809 (1981).
9. Masaki, T. & Yoshizaki, C. *J. Biochem. Tokyo* **76,** 123–131 (1974).
10. Wydro, R. M., Nguyen, H. Y., Gubits, R. & Nadal-Ginard, B. (in preparation).
11. Nudel, U. *et al. Nucleic Acids Res.* **8,** 2133–2146 (1980).
12. Nguyen, H. T., Wydro, R. M., Gubits, R. M. & Nadal-Ginard, B. *Proc. natn. Acad. Sci. U.S.A.* (submitted).
13. Umeda, P. K. *et al. Proc. natn. Acad. Sci. U.S.A.* **78,** 2843–2847 (1981).
14. Gauthier, G. F. & Lowey, S. *J. Cell. Biol.* **81,** 10–25 (1979).
15. Hoh, J. F. Y., McGrath, P. A. & Hale, P. T. *J. molec. Cell Cardiol.* **10,** 1053–1076 (1978).
16. Flink, I. L., Radar, J. H. & Morkin, E. *J. biol. Chem.* **254,** 3105–3110 (1979).
17. Pollard, T. D. & Weihing, R. *CRC crit. Rev. Biochem.* **2,** 1–65 (1974).
18. Sartore, S., Pierobon-Bormioli, & Schiaffino, S. *Nature* **274,** 82–83 (1978).
19. Lompre, A. M. *et al. Devl Biol.* **84,** 286–291 (1981).
20. Schwartz, K., Lompre, A. M., Bouveret, P., Wisnewsky, C. & Swynghedauw, B. *Eur. J. Biochem.* **104,** 341–346 (1980).
21. Hoh, J. F. Y., Yeoh, G. P. S., Thomas, M. A. W. & Higginbottom, L. *FEBS Lett.* **97,** 330–334 (1979).
22. Lompre, A. M. *et al. Nature* **282,** 105–107 (1979).
23. Mercadier, J. J. *et al. Circulation Res.* **49,** 525–532 (1981).
24. Medford, R. M., Wydro, R. M., Nguyen, H. T. & Nadal-Ginard, B. *Proc. natn. Acad. Sci. U.S.A.* **77,** 5749–5759 (1980).
25. Grunstein, M. & Hogness, D. S. *Proc. natn. Acad. Sci. U.S.A.* **72,** 3961–3965 (1975).
26. Yaffe, D. *Proc. natn. Acad. Sci. U.S.A.* **61,** 477–483 (1968).
27. Nadal-Ginard, B. *Cell* **15,** 855–864 (1978).
28. Wydro, R. M. , Periasamy, M. & Nadal-Ginard, B. (in preparation).
29. Hornig, D. & Nadal-Ginard, B. *J. biol. Chem.* (submitted).
30. Wahl, G. M., Stern, M. & Stark, G. R. *Proc. natn. Acad. Sci. U.S.A.* **76,** 3683–3687 (1979).
31. Southern, E. M. *J. molec. Biol.* **18,** 503–517 (1975).
32. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74,** 560–564 (1977).
33. Proudfoot, N. J. & Brownlee, G. G. *Nature* **263,** 211–214 (1976).
34. Elzinga, M. & Trus, B. in *Methods in Peptide and Protein Sequence Analysis* (ed. Birr, Chr.) 213–224 (Elsevier, Amsterdam, 1980).
35. Stone, D., Sode, K. J., Johnson, P. & Smillie, L. B. *Proc. IX FEBS Meet.* **31,** 125–136 (North-Holland, Amsterdam, 1975).
36. Huxley, H. E. & Brown, H. E. *J. molec Biol.* **30,** 383–434 (1967).
37. Weatherall, D. J. & Clegg, J. B. *Cell* **16,** 467–479 (1979).
38. Law, S. W. & Dugaiczyk, A. *Nature* **291,** 201–205 (1981).
39. Banerjee, S. K. & Morkin, E. *Circulation Res.* **41,** 630–634 (1977).
40. Goodkind, M. J., Dambach, G. E., Thyrum, P. T. & Luchi, R. J. *Am. J. Physiol.* **226,** 66–72 (1974).
41. Banerjee, S. K., Flink, I. L. & Morkin, E. *Circulation Res.* **39,** 319–326 (1976).
42. Smith, H. O. & Birnstiel, M. L. *Nucleic Acids Res.* **3,** 2387–2398 (1976).
43. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65,** 499–560 (1980).
44. Soeiro, R., Birnboim, H. C. & Darnell, J. E. *J. molec. Biol.* **19,** 362–372 (1966).
45. Cox, R. A. *Meth. Enzym.* **12B,** 120–129 (1968).
46. Benoff, S. & Nadal-Ginard, B. *Proc. natn. Acad. Sci. U.S.A.* **76,** 1853–1857 (1979).
47. Thomas, P. S. *Proc. natn. Acad. Sci. U.S.A.* **77,** 5201–5205 (1980).
48. Wu, C., Bingham, P. M., Livak, K. J., Holmgren, R. & Elgin, S. C. R. *Cell* **16,** 797–806 (1979).

LETTERS TO NATURE

γ-ray sources as comptonized X-ray sources

E. E. Fenimore, R. W. Klebesadel, J. G. Laros & R. E. Stockdale

University of California, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

S. R. Kane

Space Science Laboratory, University of California, Berkeley, California 94720, USA

γ-ray burst spectra have often been fit by optically thin thermal bremsstrahlung. However, at the high temperatures implied by such fits ($kT \sim 300$ keV), the free–free cross-section is so much smaller than the Compton cross-section that Compton scattering might dominate the spectral formation processes. We have investigated the possibility that emission mechanisms based on Compton scattering can also fit the data. In particular, Monte Carlo calculations have been used to compare the γ-ray burst spectral data with black-body spectra which have undergone inverse comptonization by a much hotter, overlying plasma. The best-fit parameters suggest that the underlying black body is an X-ray source ($kT_{BB} \sim 2.4$ keV) while the overlying, comptonizing plasma has $kT_e \sim 150$ keV and a column density of 4×10^{24} electrons cm^{-2}. Such a model would also naturally explain some of the unusual γ-ray burst spectra. In particular, a low energy (~400 keV) cutoff in the spectrum of GB781119 can be explained as a Wien peak produced by a higher column density in the comptonizing region, and two-component spectra (for example, GB790329) can be explained by a lower column density.

Recently, several γ-ray burst spectra have been measured with sufficient energy range and resolution to determine the continuum shape[1]. The continuum shape has usually been attributed to optically thin thermal bremsstrahlung, consistent with earlier results for the 1972 14 May burst[2] and the Apollo burst[3]. In addition, low-energy absorption features (identified as cyclotron lines) and high-energy emission features (identified as redshifted annihilation lines) have been reported[1]. These spectral features argue strongly that γ-ray bursts originate from highly magnetized neutron stars[1].

However, there are several difficulties with the above interpretations. First, the interpretation of the continuum as optically thin thermal bremsstrahlung implies a very large emitting volume (especially if the emitting region is homogeneous and if the source is further than 10 pc) whereas the cyclotron lines imply a small volume[4]. Second, various theories[5–7] suggest that the burst originates in a small volume near the surface of a neutron star rather than from a dilute, 100-km sphere implied by thermal bremsstrahlung. Third, thermal bremsstrahlung does not always fit the observations; often the observed spectrum are so hard that no thermal bremsstrahlung spectrum can fit the data. Finally, the cyclotron lines reported to date may be instrumental artefacts[4].

We have investigated the spectral distribution emerging from a plasma at temperatures typical of γ-ray burst spectra ($kT >$ 100 keV) but where there is an underlying source of photons which are reprocessed by inverse Compton scattering. As the free–free cross-section is much smaller than the Compton cross-section, the overlying plasma will have negligible thermal emission. We use a Monte Carlo method[8,9] of calculation rather than the Fokker–Planck method[10–12] because the Fokker–Planck equations break down when the temperature is >100 keV and/or when there are large-changes in energy occurring in only a few scatterings[13]. The physics of our model is similar to the two-temperature accretion disk model[14] suggested for Cygnus X-1. The primary difference from previous (X-ray) Monte Carlo calculations is that we treat rigorously the coupling between the maxwellian electron distribution and the Compton cross-section to ensure that both high-temperature effects and the angular distribution are accounted for. The angular distribution of the electron involved in the scattering can have deviation from isotropy due to the angular dependence of the cross-section.

To include temperature and angular effects, the Compton cross-section must be weighted by an isotropic maxwellian electron distribution. The total Compton cross-section for a photon of energy E_p into a plasma of temperature T_e and unit density is

$$\sigma(E_p, T_e) = 2\pi \int_0^\infty \int_0^\pi N(E_e, T_e)\sigma(E_p, E_e, \alpha)\sin\alpha \, d\alpha \, dE_e \quad (1)$$

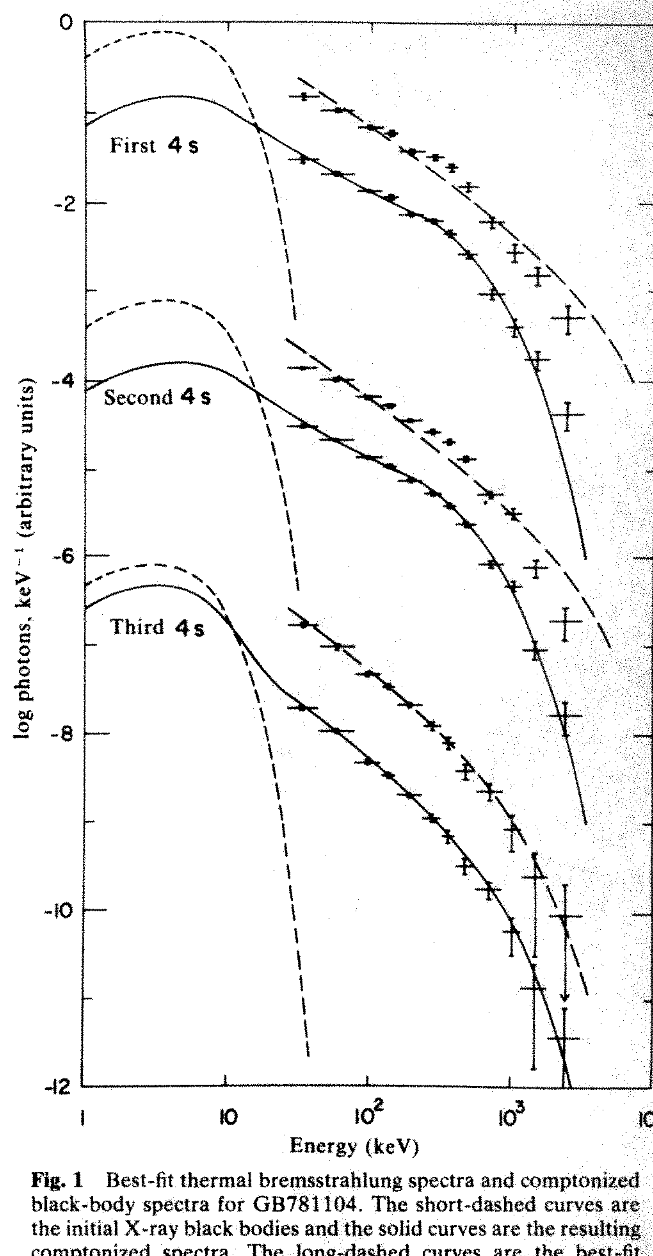


Fig. 1 Best-fit thermal bremsstrahlung spectra and comptonized black-body spectra for GB781104. The short-dashed curves are the initial X-ray black bodies and the solid curves are the resulting comptonized spectra. The long-dashed curves are the best-fit thermal bremsstrahlung spectra. The data points (repeated for the thermal bremsstrahlung and comptonized black-body fits) are from the Berkeley/Los Alamos ISEE 3 solar X-ray spectrometer.

0028-0836/82/250665—03$01.00

where the relativistic maxwellian is

$$N(E_e, T_e)\,dE_e \propto (E_e + E_0)(E_e^2 + 2E_0E_e)^{1/2} \exp(-E_e/kT_e)\,dE_e \quad (2)$$

with $E_0 = 511$ keV and $\sigma(E_p, E_e, \alpha)$ is the Compton cross-section for a photon of energy E_p onto the electrons of kinetic energy E_e moving at an angle of α with respect to the photon. To determine $\sigma(E_p, E_e, \alpha)$, a Lorentz transformation is made to a frame in which the electrons are at rest and then the cross-section is found by the Klein–Nishina formula[8]. That is, $\sigma(E_p, E_e, \alpha) = (1-\beta\cos\alpha)\sigma_{KN}[E_p\gamma(1-\beta\cos\alpha)]$ where $\gamma = 1 + E_e/E_0$, $\beta = (1-\gamma^{-2})^{1/2}$ and σ_{KN} is the Klein–Nishina cross-section.

For this preliminary study, we assumed that the emitting region consisted of a point source of black-body radiation at the centre of a spherical volume containing a hot thermal plasma with no magnetic field. Three parameters were chosen to characterize the region: T_{BB}, the black-body temperature of the central source; ρx, the column density (that is, the radius times the electron density); and T_e, the temperature of the comptonizing volume. The following steps were taken to track each photon:

(1) A random number, r_1, between 0.0 and 1.0 is selected and used to determine the distance the photon travels before undergoing a Compton interaction:

$$d = -(N_0\sigma(E_p, T_e))^{-1}\ln(r_1) \quad (3)$$

where N_0 is the electron density of the plasma. If the location of the proposed Compton scattering site is outside the assumed size of the volume, the photon is considered to have escaped and the emerging energy spectrum is incremented at the photon's energy.

(2) Once a location for the Compton scattering is known, the energy of the electron involved is found by solving for E_e in

$$r_2 = \frac{1}{\sigma(E_p, T_e)}\int_0^{E_e}\int_0^{\pi} N(E, T_e)\sigma(E_p, E, \alpha)\sin\alpha\,d\alpha\,dE \quad (4)$$

where r_2 is a new random number.

(3) The angle between the photon and the electron is found by solving for α_e in

$$r_3 = \int_0^{\alpha_e}\sigma(E_p, E_e, \alpha)\sin\alpha\,d\alpha \Big/ \int_0^{\pi}\sigma(E_p, E_e, \alpha)\sin\alpha\,d\alpha \quad (5)$$

The other angle to complete the definition of the photon's direction relative to the electron's is found simply as $2\pi r_4$.

(4) Knowing E_p, E_e, and α_e allows a Lorentz transformation to the rest frame of the electron where the Compton interaction is calculated. Two additional random numbers, coupled with the formulae for the radiation pattern of Compton scattering on an electron at rest, give the scattering angle and the change in energy of the photon in the electron rest frame. The inverse Lorentz transformation gives the new energy and direction of the photon after the Compton collision in the laboratory (that is real world) frame. Steps (1) to (4) are repeated until the photon escapes.

The results of the Monte Carlo calculations have been compared directly with the observed spectrum of the typical and well studied burst GB781104 (refs 1, 4, 15, 16). It was not computationally feasible to perform a Monte Carlo calculation at each combination of T_e and ρx required to determine the best-fit parameters and their errors. Rather, calculations were performed on a grid of T_e and ρx values which were then interpolated to obtain a spectrum for a particular set of T_e and ρx. Table 1 gives the best fit for both thermal bremsstrahlung and comptonized black bodies with their joint 68% confidence levels (calculated by the method of ref. 17) for three successive 4-s intervals. In Table 1, χ_c^2, N_{df}, T_{TB}, χ_{TB}^2 are, respectively, the χ^2 of the comptonized black body fit. the number of degrees of freedom, the best-fit thermal bremsstrahlung temperature, and the χ^2 for the thermal bremsstrahlung fit. Figure 1 shows, for each 4 s interval, the initial black body as a short-dashed curve, the resulting comptonized spectrum as a solid curve with ISEE 3 solar X-ray spectrometer data (with 1 σ error bars), and the best-fit thermal bremsstrahlung spectrum (long-dashed curve) with the ISEE 3 data. The reconstructed spectrum obtained by unfolding the observed counts depends on the assumed spectral shape. For this reason, the unfolded ISEE 3 data points differ slightly depending on whether thermal bremsstrahlung or a comptonized black body is assumed. For clarity, the thermal bremsstrahlung fit (with unfolded ISEE 3 data points) is offset from the comptonized black body fit (and its data) in Fig. 1. The optically thin thermal bremsstrahlung formula included the relativistic Guant factor of ref. 18.

Optically thin thermal bremsstrahlung could not satisfactorily fit the data during either of the first two 4-s intervals due to the relative hardness of the spectrum. Only extremely high temperature thermal bremsstrahlung came close to fitting the data. In contrast, the comptonized spectrum had a much more reasonable plasma temperature ($T_e = 155$ keV compared with $T_{TB} = 3{,}800$ keV) and a much lower value of χ^2 (for example, 19.0 for the comptonized black body compared with 140.0 for thermal bremsstrahlung). (The thermal bremsstrahlung fit is not substantially improved even if a broad emission line between 230 and 600 keV is assumed; T_{TB} is reduced to 3,300 keV with $\chi_{TB}^2 \sim 80$.) Thermal bremsstrahlung would not fit our data in the second interval either, giving a χ_{TB}^2 of 330 whereas the comptonization model gives a χ_C^2 of 18.6. These values of χ_C^2, although much better than those for thermal bremsstrahlung, are marginally too high to be considered statistically acceptable. However, the spectrum is changing on a time scale shorter than our sampling rate (4 s), which might explain these rather large values of χ^2. The third 4-s interval could be fit by either a thermal bremsstrahlung spectrum (with $\chi_{TB}^2 = 10.7$) or a comptonized black body (with $\chi_C^2 = 8.8$).

Note that, in all three intervals, the best-fit black body has an X-ray temperature of ~ 2 keV. The χ^2 surfaces are not quadratic functions of the parameters and, thus, one cannot estimate from the 68% confidence level the parameter limits for other confidence levels. The term in Table 1, $kT_{BB,95\%}$, gives the lowest black body temperature that fits the data at the 95% confidence level. Although the first 4-s interval could be consistent with a very cool source of photons, the next two intervals are only consistent with black-body temperatures >0.4 keV. There is no evidence to indicate that the underlying X-ray source is any known type of X-ray activity, in particular, a burster. Rather, the X-ray source is probably a result of heating of the surface by the overlying, hotter region.

One potential difficulty with our model is that the cooling time of the overlying plasma is $\sim 10^{-6}$ s, much shorter than the duration of the bursts. Thus, some process must replenish the overlying region. A similar problem is also found in most other γ-ray burst models[7].

Three of the assumptions we have made are: (1) the source is at the centre of a sphere; (2) the emerging spectrum can be found by interpolation between Monte Carlo calculations; and

Table 1 Best-fit parameters for GB781104

	First 4 s	Second 4 s	Third 4 s
$\chi_{TB}^2 (N_{df} = 10)$	140	330	10.7
kT_{TB} (keV)	3,800	2,334	786±70
$\chi_C^2 (N_{df} = 8)$	19.0	18.6	8.8
kT_e (keV)	155±35	160±20	$240^{+\infty}_{-70}$
$\rho x (10^{24}$ e cm$^{-2})$	4.5±1.2	4.0±0.6	1.2±1.2
kT_{BB} (keV)	$2.4^{+9.6}_{-2.4}$	2.4±1.5	2.0±0.4
$kT_{BB,95\%}$ (keV)	0.0	0.4	0.7

(3) there is no magnetic field. Preliminary calculations indicate that the first two assumptions do not cause large changes in the shape of the resulting spectrum, although changes in the best-fit parameters could occur. Recently, very high magnetic fields ($\sim 10^{12}$ G) have been assumed in γ-ray burst models[5-7]. However, we have made a preliminary comparison of the spectral shapes predicted for cyclotron processes (see ref. 7) with γ-ray burst spectra. Cyclotron spectra seem to be much softer than typical γ-ray burst spectra. A very hard spectrum (such as GB781104) would be even more difficult to fit. To prevent cyclotron processes from dominating the radiation transport, the magnetic field in the comptonizing region must be $\lesssim 10^9$ G. Even a field as low as 10^9 G would be able to confine the plasma on the time scale of replenishment.

There are two other features of some γ-ray burst observations which could naturally be explained by comptonized X-ray models. In some bursts, the spectrum consists of a normal γ-ray component plus a soft component (with $kT < 20$ keV)[1]. The soft component is discernible during periods when the γ-ray flux above 100 keV has dropped to near background. Two examples of such bursts are GB790329 and GB790524 (see Fig. 2 in ref. 1). The Comptonized X-ray burst model would give a similar two-component spectrum when the density of the overlying region decreases so that some low-energy photons escape without having scattered, and thus, retain the spectrum of the X-ray black body. Those photons that do scatter produce the normally observed γ-ray component.

The second observation which might be explained by the comptonized X-ray model concerns the γ-ray burst of 19 November 1978. The spectrum initially had a low energy cutoff starting at $\sim$400 keV; as the spectrum evolved, the cutoff moved to lower energies resulting in a monotonic-appearing spectrum[19,20]. This behaviour is consistent with a comptonized X-ray model if the density of the scattering medium was high. At high densisties, there is a sufficient number of scatterings to equilibrate the photons, thus producing a Wien peak. If the density decreases, the spectrum would become monotonic.

Most of the annihilation lines[1] are probably too narrow to be explained by Wien peaks. In fact, the broad emission feature in 19 November 1978 was not identified as an annihilation line by Mazets *et al.*[1] Mazets *et al.*[1] claim an annihilation line in 4 November 1978 whereas our comptonized black body fits do not require such a line. However, a smooth line through our data points in the thermal bremsstrahlung unfolding during the first 4 s would seem to require a line at about the intensity and shape reported by Mazets *et al.*[1] (see Fig. 1). This demonstrates how spurious line features can be produced if a wrong continuum shape is assumed.

The best-fit parameters indicate that γ-ray burst continua are consistent with an X-ray source which undergoes inverse comptonization by a hot overlying plasma. Such a hot plasma is normally assumed for a γ-ray burst and we offer no explanation of why the hot plasma would be overlying an X-ray source. Rather, these results indicate that a particular continuum shape other than thermal bremsstrahlung can sometimes better fit the available data. The implications of these fits should be explored with more detailed theoretical work.

We thank A. G. Petschek, F. A. Córdova and W. C. Priedhorsky for helpful discussions and comments. The Berkeley/Los Alamos ISEE 3 solar X-ray spectrometer has been supported in part by NASA under contract NAS5-25980. This work was done under the auspices of the US Department of Energy.

Received 7 January; accepted 22 April 1982.

1. Mazets, E. P., Golenetskii, S. V., Aptekar, R. L., Guryan, Y. A. & Ilyinskii, V. N. *Nature* **290**, 378 (1981).
2. Wheaton, W. A. *et al. Astrophys. J. Lett.* **185**, L57 (1973).
3. Gilman, D., Metzger, A. E., Parker, R. H., Evans, L. G. & Trombka, J. I. *Astrophys. J.* **236**, 951 (1980).
4. Fenimore, E. E., Laros, J. G., Klebesadel, R. W., Stockdale, R. E. & Kane, S. R. *Proc. Workshop on Gamma Ray Bursts and Related Transient Phenomenon*, UCSD (1981).
5. Woolsey, S. E. *Proc. Workshop on Gamma Ray Bursts and Related Transient Phenomenon*, UCSD (1981).
6. Ramaty, R., Bonazzola, S., Cline, T. L., Kazanas, D. & Mesyaros, P. *Nature* **287**, 122 (1980).
7. Lamb, D. *Proc. of Workshop on Gamma Ray Bursts and Related Transient Phenomenon*, UCSD (1981).
8. Loh, E. D. & Garmire, G. P. *Astrophys. J.* **166**, 301 (1971).
9. Pozdnyakov, L. A., Sobol, I. M. & Syunyaev, R. A. *Soviet Astr. Lett.* **2**, 55 (1976).
10. Kompaneets, A. S., *Soviet Phys.-JETP* **4**, 730 (1957).
11. Illarionov, A. F. & Sunyaev, R. A. *Soviet Astr.* **16**, 45 (1972).
12. Sunyaev, R. A. & Titarchuk, L. G. *Astr. Astrophys.* **86**, 121 (1980).
13. Langer, S. H. *Astrophys. J.* **232**, 891 (1979).
14. Shapiro, S. L., Lightman, A. P. & Eardley, D. M. *Astrophys. J.* **204**, 187 (1976).
15. Evans, W. D. Fenimore, E. E., Klevesadel, R. W., Laros, J. G. & Terrell, N. J. *Astrophys. Space Sci.* **75**, 35 (1981).
16. Teegarden, B. J. & Cline, T. L. *Astrophys. J. Lett.* **236**, L67 (1980).
17. Lampton, M., Margon, B. & Bowyer, S. *Astrophys. J.* **208**, 177 (1976).
18. Matteson, J. L. thesis, Univ. California, San Diego (1972).
19. Barat *et al. Paper 23rd COSPAR Meet.*, Budapest (1980).
20. Vedrenne, G. *Phil. Trans. R. Soc.* **A301**, 645 (1981).

Metal–organic coating interface as studied by emission Mössbauer spectroscopy

Henry Leidheiser Jr, Svetozar Musić* & Gary W. Simmons

Center for Surface and Coatings Research, Lehigh University, Bethlehem, Pennsylvania 18015, USA

A major unknown in an understanding of the interface between an organic coating and a metal substrate is the chemical nature of the interfacial region. This is specially important in understanding the phenomena that control 'wet adhesion'[1], the ability of coated metals to withstand long exposure to high humidity conditions, and the disbonding of light-sensitive polymers used in lithographic plates and photoresists. We report here how ^{57}Co emission Mössbauer spectroscopy can be used to study the chemistry of the interfacial region in a model system consisting of cobalt metal and a commercial polybutadiene coating.

A previous study showed that the Mössbauer technique was suitable for studying corrosion beneath a coating[2]. Cobalt is selected here as the substrate metal because the isotope ^{57}Co is Mössbauer-active and can be concentrated at the metal surface using electrodeposition. Polybutadiene is selected as the coating because of our previous work with this material[3] and the utilization of polybutadiene by Dickie and co-workers[4] and by Castle and Watts[5] in corrosion studies.

Emission Mössbauer spectroscopy is based on the 14.4-keV γ ray that is emitted from ^{57}Fe as a consequence of the electron capture decay of ^{57}Co. Although the Mössbauer spectra obtained with this method are of ^{57}Fe, it is likely that it will be possible to obtain information about the chemical state of the initial ^{57}Co isotope. Indeed, in many cases emission spectra of cobalt compounds give hyperfine parameters expected on the basis of known oxidation state, structure and magnetic properties. It is also commonly observed that a fraction of the oxidation states of ^{57}Fe is atypical of the original oxidation state of ^{57}Co. These atypical oxidation states are produced as a consequence of the Auger cascade that follows the electron capture decay of ^{57}Co (refs 6–8). The fraction of these states that are stabilized is a function of chemical environment. For example, no atypical ^{57}Fe oxidation states are found in metallic environments, whereas in other compounds, such as hydrated cobaltous salts, a significant fraction of Fe^{3+} is found. Our interpretation of the emission Mössbauer data is based on extensive emission studies of the oxides, hydroxides and hydroxyoxides of cobalt that may be present at a cobalt metal–polymer interface[9-11]. The reported changes in the emission spectra of cobalt at this interface represent either direct changes in the chemical state of cobalt or changes in the environment of cobalt that in turn affects the stabilization of atypical ^{57}Fe charge states.

Cobalt, doped with ^{57}Co, was electrodeposited at room temperature on a cobalt substrate at a current density of 0.01 A cm^{-2} for 5–20 s while the metal was immersed in a bath

* Permanent address: Rudjer Bošković Institute, Zagreb, Yugoslavia.

Table 1 Mössbauer spectra parameters

Sample spectrum of electrodeposit	Component	δ (mm s^{-1})	Δ (mm s^{-1})	$H_{5/2}$ (kOe)	Γ (mm s^{-1})	Area of component (mm s^{-1})	A (%)
Before applying the polybutadiene coating (Fig. 1)	Fe^{3+}	−0.093	0.808		0.996	0.284	71
	Fe^{2+}	−0.952	2.248		0.540	0.016	4
	Fe^{0}	0.398	0.124	338	0.885	0.097	24
After applying polybutadiene coating but before baking	Fe^{3+}	−0.149	0.792		1.026	0.277	67
	Fe^{2+}	−0.925	2.096		0.786	0.020	5
	Fe^{0}	0.532	0.168	336	1.241	0.114	28
After baking the polybutadiene coating for 30 min at 200 °C (Fig. 2)	Fe^{3+}	−0.127	0.701		1.032	0.155	38
	Fe^{2+}	−0.883	2.214		1.216	0.153	38
	Fe^{0}	0.380	0.220	337	1.088	0.096	24

Chemical shift (δ) is given with reference to an enriched $K_4Fe(CN)_6$ absorber; Δ = quadrupole splitting value; $H_{5/2}$ = hyperfine magnetic splitting. Γ = line width; A = relative amount of component.

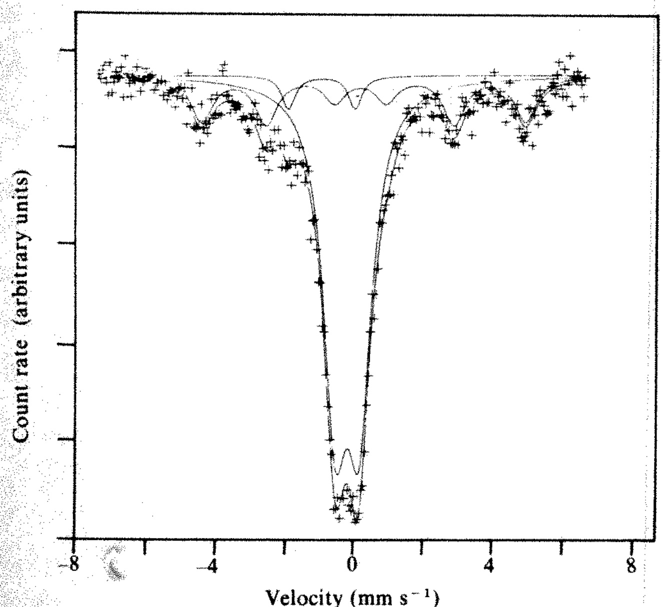


Fig. 1 ^{57}Co Mössbauer emission spectrum of a thin cobalt electrodeposit used as the starting point in the experiment described in the text.

containing 0.31 g l^{-1} $CoSO_4{\cdot}7H_2O$, 100 g l^{-1} $MgSO_4{\cdot}7H_2O$, and 30 g l^{-1} H_3BO_3, pH = 4.0. The calculated thickness of the deposit was 20–120 nm assuming a current efficiency of 100%, but the actual thickness was considerably less. The Mössbauer spectrum of a thin deposit is shown in Fig. 1 in which the metallic component (six-line spectrum) was ~24% and the ionic component was ~76%. These values are calculated on the basis of the area under the curve for each component and assume that the recoilless fraction is the same for all components. The ionic component presumably results from superficial oxidation of the metallic surface on exposure to air. The ratio of the metallic to the ionic component increased as the thickness of the electrodeposit was increased. Maximum sensitivity in the experiments to be described was obtained with deposits which yielded a spectrum similar to that in Fig. 1. The major doublet is at a position characteristic of Fe^{3+} and only a small amount of Fe^{2+} (4%) is revealed by computer curve fitting. On the basis of the studies of Barr[12] and of Rice *et al.*[13], it is likely that the outer layer of cobalt exposed to the atmosphere has a composition equivalent to CoOOH. Identification of the surface composition by Mössbauer techniques alone does not appear practical as $Co(OH)_2$ exhibits spectra showing different amounts of Fe^{2+} and Fe^{3+} depending on the method of preparation[1].

The cobalt surface was then coated with a commercial polybutadiene coating (Budium RK-662, DuPont) to a thickness of ~15 μm and the emission spectrum was again obtained without baking the coating. The spectrum was very similar to that shown in Fig. 1. The Mössbauer parameters are given in Table 1. The coated metal was cured by baking at 200 °C in air for 30 min and a spectrum was again obtained. As shown in Fig. 2 and in Table 1, the Fe^{2+} doublet is now a substantial part of the total spectrum. Two possible changes at the interface could account for the increase in the Fe^{2+} component. First, an interaction between the cobalt oxide and the polymer may have reduced Co^{3+} in the surface oxide to Co^{2+}. Second, the treatment may have led to the dehydration of a Co^{2+} oxide with the result of destabilization of Fe^{3+}. (Stabilization of Fe^{3+} in the presence of water is thought to be the result of oxidation of Fe^{2+} to Fe^{3+} by OH radicals produced by the radiolysis of water molecules by the Auger electron ionization[6–8].) The reduction of the surface oxide is supported by the tentative conclusion of Castle and Watts[5] that the oxide on iron is partially reduced during the baking of polybutadiene coating on steel. The latter conclusion was tentative since it was based on X-ray photoelectron sputtering studies of stripped oxides and it is well known that the sputtering process may lead to changes in valence states of the cations in oxides[14].

The experiments described in Table 1 were repeated several times with similar results showing that the Mössbauer technique is sensitive at the level required to give information about the chemical nature of the organic coating–metal interface. The following conclusions can be drawn: (1) The contact between the cobalt metal and the polybutadiene is through an oxide or hydrous oxide at the interface. (2) No major changes at the cobalt interface occurred after application of the coating and

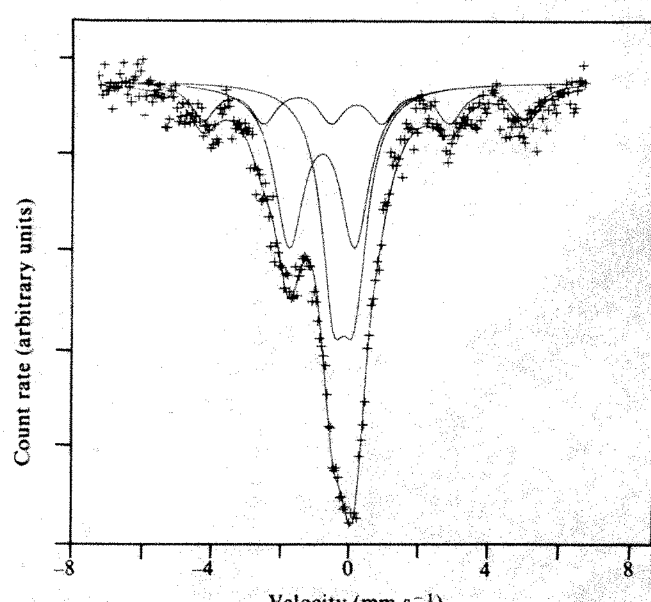


Fig. 2 ^{57}Co Mössbauer emission spectrum of sample after coating with polybutadiene and baking at 200 °C for 30 min. Compare spectrum with that in Fig. 1.

before thermal treatment. (3) The oxide on cobalt does not thicken during the baking process. (4) Thermal treatment at 200 °C causes a significant amount of conversion of Fe^{3+} to Fe^{2+} at the oxide–coating interface. (5) The reduction process cannot be a consequence of the reaction, $2Co^{3+} + Co = 3Co^{2+}$, during baking of the coating because the relative amount of unoxidized metal remains approximately constant before and after baking. (6) The conversion of Fe^{3+} to Fe^{2+} may be attributed to either a reduction of Co^{3+} to Co^{2+} or to the destabilization of Fe^{3+} by the dehydration of Co^{2+} oxide.

Small changes in the chemical shift and in the quadrupole splitting values before and after baking the polybutadiene may be significant, but we prefer to wait for further experiments with other metal–polymer coating systems before drawing further conclusions.

We thank the Office of Naval Research for support.

Received 2 March; accepted 4 May 1982.

1. Funke, W. *Corrosion Control by Coatings* (ed. Leidheiser, H. Jr) 35–45 (Science Press, Princeton, New Jersey, 1979).
2. Leidheiser, H. Jr, Simmons, G. W. & Kellerman, E. *J. electrochem. Soc.* **120**, 1516–1518 (1973).
3. Leidheiser, H. Jr & Wang, W. *J. Coatings Technol.* **53**, No. 672, 77–84 (1981).
4. Dickie, R A., Hammond, J. S. & Holubka, J. W. *Ind. Eng. Chem. Prod. Res. Dev.* **20**, 339–343 (1981).
5. Castle, J. E. & Watts, J. F. *Corrosion Control by Organic Coatings* (ed. Leidheiser, H. Jr) 78–86 (National Association Corrosion Engineers., Houston, 1981).
6. Wertheim, G. K. *Accts Chem. Res.* **4**, 373–379 (1971).
7. Greenwood, N. N. & Gibb, T. C. *Mössbauer Spectrscopy*, 373–379 (Chapman and Hall, London, 1971).
8. Freidt, J. M. & Danon, J. *Radiochim Acta* **17**, 173–190 (1972).
9. Simmons, G. W. Kellerman, E. & Leidheiser, H. Jr. *J. electrochem. Soc.* **123**, 1276–1284 (1976).
10. Simmons, G. W. & Leidheiser, H. Jr. *Applications of Mössbauer Spectroscopy* Vol. I (ed. Cohen, R. L.) 85–125 (Academic, New York, 1976).
11. Simmons, G. W. *Proc. 4th int. Symp. on Passivity* 898–917 (The Electrochemical Society, Princeton, 1978).
12. Barr, T. L. *J. phys. Chem* **82**, 1801–1810 (1978).
13. Rice, D. W., Phipps, P. B. P. & Temoureux, R. *J. electrochem. Soc.* **126**, 1459–1466 (1979).
14. Kelly, R. *Surface Sci.* **100**, 85–107 (1980).

Tabular icebergs in ocean waves

Monica Kristensen, Vernon A. Squire & Stuart C. Moore

Scott Polar Research Institute, University of Cambridge, Cambridge CB2 1ER, UK

Ocean waves are a principal agent in the deterioration and ultimate breakup of Antarctic tabular icebergs in the Southern Ocean. We present here some preliminary results from a recent (January 1982) field season on board HMS *Endurance* during which two tabular icebergs of very different shape were studied in detail. The data augment the results of a similar season of work[1,2] which took place in 1981 near the South Sandwich and South Orkney Islands. Instruments to measure rigid body motions and strain gauges to determine wave-induced bending were deployed at the surface centre of the two icebergs visited, while a Waverider buoy simultaneously monitored the ocean wave-energy spectrum. The precise underwater shape of each iceberg was determined by flying a 60-MHz radar (developed for use over land ice)[3,4] in a grid pattern across the surface of the iceberg. By this means a complete three-dimensional picture of the iceberg could be inferred. The results show that icebergs tend to act as low-pass filters and inhibit short period waves. Furthermore, they selectively resonate at certain wave periods; the strain data, in particular, indicate unexpectedly large flexure which cannot be explained by simple bending alone. Finally, the geometry of icebergs can be such as to render them unstable and liable to turn over.

The first iceberg, which was encountered close to Clarence Island in the South Shetlands, was about 1,200 m long by 480 m wide with a freeboard varying between 39 and 57 m. Radio echo profiles showed the iceberg to be roughly trapezoidal in cross-section with the bottom sloping along its length. The second iceberg was located further north at ~62°51.3'S, 53°31.7'W. It was much larger (3,630 × 1,740 m), but had a freeboard of only 20–28 m. The underside was approximately flat with no evidence of bottom crevassing. Figure 1 represents a longitudinal radio echo transect for this iceberg showing the position of the surface and bottom echoes. A feature of Fig. 1, which has been found in many of our radar profiles, is the region of weak bottom return echo to the left of the centre of the figure. This phenomenon is possibly due to a refrozen crevasse where the new ice contains seawater salts of relatively higher dielectric absorption.

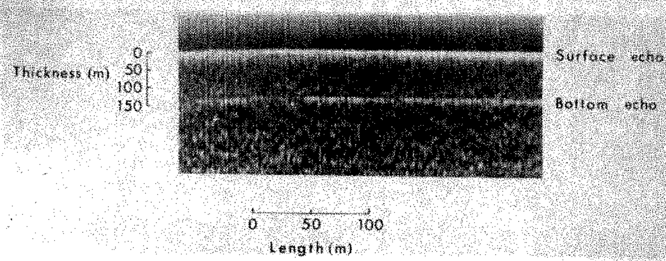


Fig. 1 Radio echo sounding profile showing upper and lower surfaces and calculated scales.

The overall experimental plan benefited considerably from the experiment which had taken place the previous year. Two Royal Navy Wasp helicopters were used to ferry instruments and personnel onto the surface of the ice, and to carry out the radio echo sounding. On the surface, three accelerometers mounted on gimbals were used to measure heave, surge and sway, and two tiltmeters measured pitch and roll. Three 1-m deep trenches in a 120° rosette configuration[5,6] were dug for the strain measuring equipment. The three strainmeters had been developed by the Scott Polar Research Institute (SPRI) specifically for use on sea ice[7]. Their use on an iceberg surface presented serious mounting problems as the top 30 m are made up of firn and snow interspersed with ice layers. A satisfactory mount was finally achieved by driving 2-m long aluminium cylinders into the surface and then rigidly fixing the strainmeters to them. Simultaneous measurements of the wave field in the waters around the berg were collected using a Waverider buoy located several kilometres away.

The main features of the icebergs' motions can be inferred from Fig. 2, where the original time series, and the unsmoothed and smoothed power spectra are presented for the larger iceberg. As might be expected, the bandwidth of the Waverider buoy data is much greater than that of any of the sensors located on the iceberg itself. This demonstrates that icebergs act to filter out the shorter wave periods found in the forcing spectrum. The 8-s cutoff is very distinct for surge, roll and strain, though the iceberg shows energy in heave at periods as low as 6 s. It is also significant that the dominant peak in the Waverider spectrum does not necessarily occur at the same period as those in the other spectra. In heave the peak occurs at about 16 s, in surge at 10 s, and in roll at 11 s. This agrees with rough calculations to find the respective natural frequencies, and indicates that the iceberg can tune to the incident ocean waves and resonate at a period which is dependent on its geometry and structure. This is especially apparent in the strain spectrum which shows three distinct peaks at periods of 10, 15 and 50 s, showing that the iceberg is extracting energy from the ocean waves at selective periods. The strains corresponding to these peaks are of the order of 10^{-6}, considerably greater than we would expect unless a resonance was occurring (10^{-6} is typical of sea ice floes two orders of magnitude thinner[8]).

Figure 3, showing another startling example of resonance seen on the first iceberg, is a copy of the roll data as recorded on the iceberg during the experiment. Dominant periods in the data are clear at about 14 and 100 s. The shorter of these oscillations represents the ocean wave forcing, the longer is the natural period of roll for an iceberg of this shape and structure. It is interesting to speculate about the origin of the 100-s roll. The oscillation might have been generated impulsively or be continually maintained by some energy in the open water at

0028-0836/82/250669—03$01.00

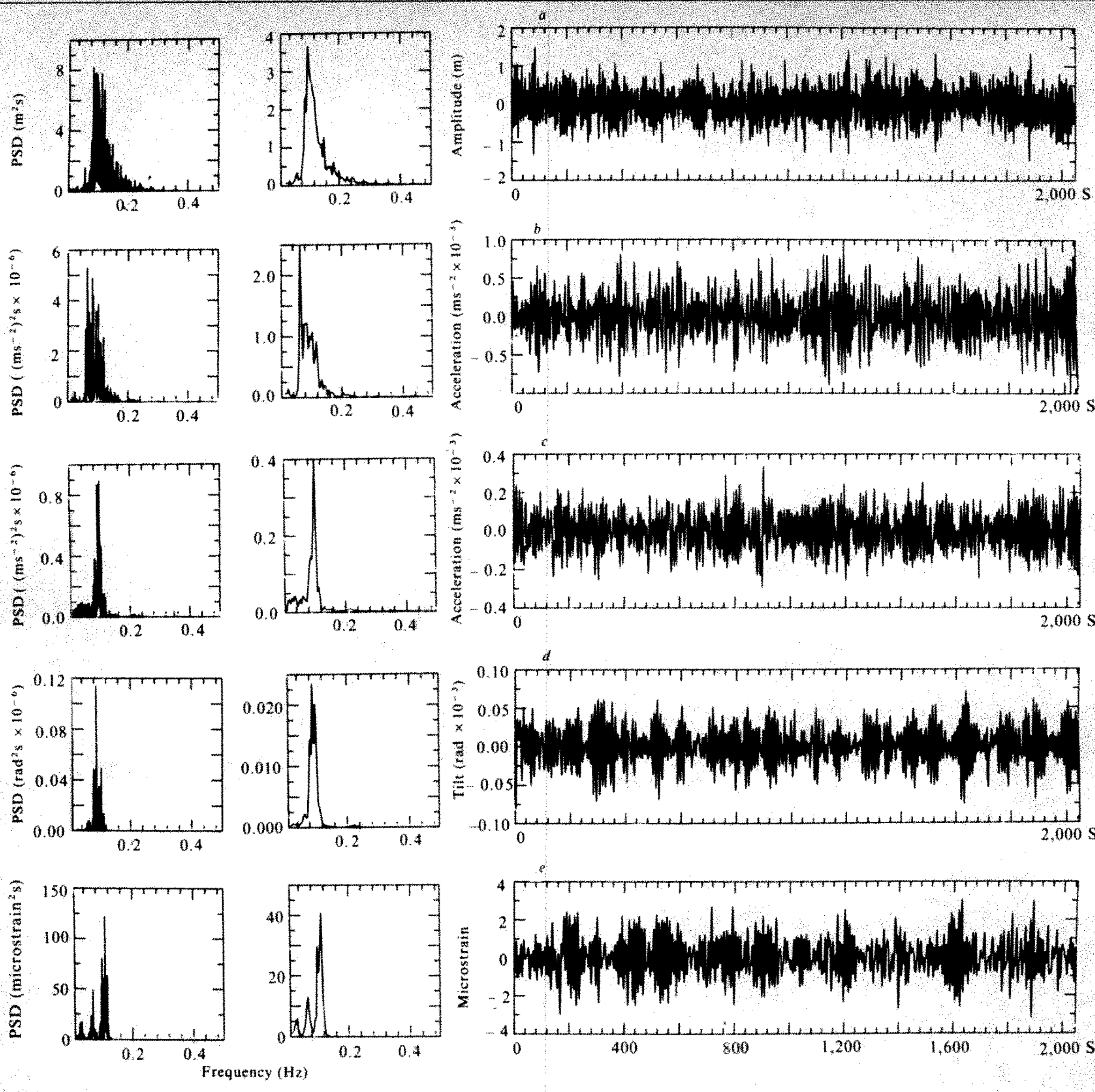


Fig. 2 Time series and computed power spectra for *a*, the Waverider buoy data and *b*, the heave; *c*, surge; *d*, roll; and *e*, strain motions of a tabular iceberg. The smoothed spectra are grouped by nine continuous power spectral density (PSD) values.

that period. In the first case the motion would decay at a rate determined by the '*Q*' of the ice–water system. In the second, the energy might come from wave packets forcing fluid motions on the length scale of the modulation envelope.

A simple calculation drawn from Lewis' review on the motion of ships in waves[9] may be used to interpret the 100-s roll. Figure 4 gives the necessary parameters for the calculation. The metacentric height, *GM*, for the iceberg is given by

$$GM = \frac{a^2}{3d} - l + \frac{d}{2} \qquad (1)$$

The expression given by Lewis for the natural period of roll is

$$T^2 = \frac{4\pi^2 r^2}{g\,GM} \qquad (2)$$

where r is the effective radius of gryration about the axis of roll which includes a contribution from the added inertia of the fluid entrained by the rolling iceberg, and g is the acceleration due to gravity. With appropriate values for the width (480 m), the thickness (273 m) and the density (850 kg m^{-3}—assumed constant but reduced because of any density variations in the berg), the effective radius of gyration is found to be nearly 400 m. This is more than double the value for the iceberg if the fluid's contribution is neglected, and emphasizes that added inertia must be taken into account if reliable theoretical estimates of rolling period are to be found. Finally, it is of interest to investigate the stability of the rolling iceberg. Adopting the suggestion of Weeks and Mellor[10] that an iceberg should be considered unstable if its metacentric height is <10% of its width, we find that the iceberg was just stable.

In conclusion, it must be stressed that there is a singular lack of data on the behaviour of icebergs in ocean waves. Two field seasons have been carried out by SPRI in cooperation with Norsk Polarinstitutt, and Foldvik *et al.*[11] have reported some measurements which took place during the Norwegian Antarctic Research Expedition of 1978–79. The results from the two SPRI seasons together constitute a unique data set which may

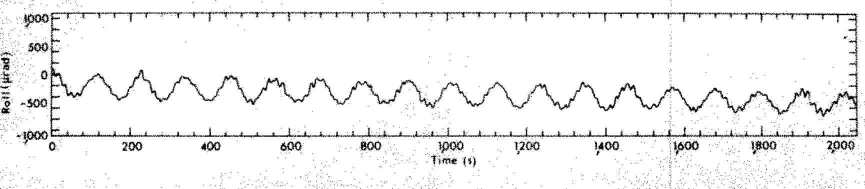


Fig. 3 Roll data record from first iceberg.

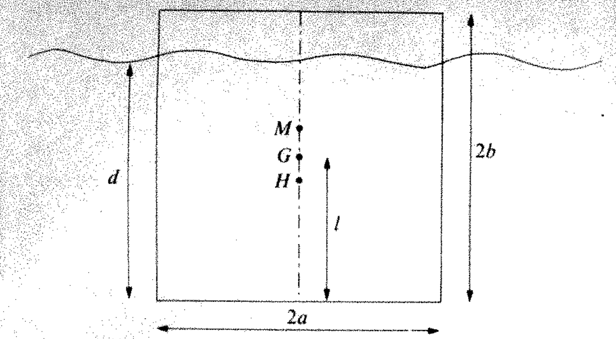


Fig. 4 Simplified rectangular iceberg section. M is the iceberg's metacentre, G is its centre of mass, and H is its centre of buoyancy (the centre of mass of the displaced fluid).

be used to predict sea-keeping characteristics for tabular icebergs generally. Foldvik's discussion is unfortunately limited to tilt alone as his strain measuring system was unable to resolve the slight cyclic bending which occurred. In presenting his tilt data, Foldvick has derived theoretical curves based on equation (2) but has neglected added inertia. We have shown that this can seriously underestimate any natural roll period calculations. The data reported here fall at the two extremes of his measurements.

We thank the hydrographer of the Royal Navy for permission to join HMS *Endurance*, and Captain Barker, the officers and the crew of the ship for their patience and help. We thank the Norsk Polarinstitutt (NP) and the Radio Echo Group at SPRI for loan of equipment, Drs Peter Wadhams and Olav Orheim for advice on experimental design based on 1981 experience, Rob Massom for diagrams, and Arne Gaard for stimulating discussion. The project was assisted financially by BP Co. Ltd. M.K. is research student supported by a British Council grant, and by Roald Amundsen Minnefond and NP field expenses.

Received 19 April; accepted 25 May 1982.

1. Kristensen, M., Orheim, O. & Wadhams, P. *Polar Rec.* **20**, 445–457 (1981).
2. Orheim, O., Wadhams, P. & Kristensen, M. *Ann. Glaciol.* **3**, 357–358 (1982); *Iceberg Res.* **1**, 10–15 (1982).
3. Evans, S. & Smith, B. M. E. *J. scient. Instrum.* **2**, 131–136 (1969).
4. Robin, G. de Q. Evans, S. & Bailey, J. T. *Phil. Trans. R. Soc.* **A265**, 437–505 (1969).
5. Jaeger, J. C. *Elasticity, Fracture and Flow with Engineering and Geological Applications* (Methuen, New York, 1956).
6. Squire, V. A. *J. Glaciol.* **29**, 425–431 (1979).
7. Moore, S. C. & Wadhams, P. *Tech. Rep.* No. 81-2 (Scott Polar Research Institute, Cambridge, 1981).
8. Squire, V. A. & Martin, S. *Sci. Rep.* No. 18 (School of Oceanography, University of Washington, Seattle, 1980).
9. Lewis, E. V. *Principles of Naval Architecture* (ed. Comstock, J. P.) 607–717 (The Society of Naval Architects and Marine Engineers, 1967).
10. Weeks, W. F. & Mellor, M. *Iceberg Utilization, Proc. of the First int. Conf., Ames, Iowa*, 45–98 (Pergamon, New York, 1978).
11. Foldvik, A., Gammelsrod, T. & Gjessing Y. *Ann. Glaciol.* **1**, 29–30 (1980).

A deep seismic reflection profile over a Caledonian granite in central England

P. K. H. Maguire*, E. M. Andrew†, G. Arter*, R. A. Chadwick‡, P. Greenwood†, I. A. Hill*, N. Kenolty‡ & M. A. Khan*

* Department of Geology, University of Leicester, Leicester LE1 7RH, UK
† Applied Geophysics Unit and ‡ Deep Geology Unit, Institute of Geological Sciences, Keyworth, Nottingham NG12 5GG, UK

In 1980, a short, deep reflection profile, undertaken jointly by the University of Leicester and the Institute of Geological Sciences, crossed a buried Caledonian granite to the south-west of Melton Mowbray in Leicestershire. Little structure is seen in the section between the top of the granite at 0.25 s two-way time (s TWT), and a zone of very strong sub-horizontal reflections at 6 s TWT. A further reflection zone is identified at 7.2 s TWT below which there are a series of strong but impersistent reflections to the full reording time of 12 s TWT. The data indicate that the granite must be relatively homogeneous and underlain by several marked sub-horizontal boundaries at mid- to lower-crustal levels. This suggests that, during emplacement, either the granite must have had little effect on the lower crust sampled by the profile, or the lower crust was mobile enough to have developed layering since the granite was emplaced during the Caledonian orogeny.

The various granitic rocks of Leicestershire observed at Mountsorrel, Countesthorpe, Croft, Enderby, Stoney Stanton and in the recent NCB borehole at Kirby Lane, south-west of Melton Mowbray (Fig. 1) are considered to be the surface expression of a large calc-alkaline granitic pluton[1,2]. These granitic rocks generally have a strong magnetic signature and have been shown to be distinguishable from the surrounding Precambrian, meta-sedimentary and meta-volcanic rocks at basement level by their near surface, lower P-wave seismic velocity, derived from seismic refraction data[3].

A particularly suitable problem for a limited-budget, deep seismic reflection programme was to examine whether a granite may provide a transparent medium for acoustic energy to penetrate the lower crust. It was therefore decided that a survey should be undertaken over the large, 250 nT, aeromagnetic anomaly associated with the Melton Mowbray granite, in particular along a disused railway line passing within 1 km of the Kirby Lane borehole.

Data acquisition was carried out using the following parameters: 9 kg explosive charges at 15 m depth, recorded by a 24 channel 'end-on' array: 12 geophones per station at a 50 m station interval. Shooting every station resulted in 12-fold multiplicity, horizontal subsurface cover being 2.125 km. Data were recorded to 12 s TWT and processed at a 4-ms sample rate, using a routine oil industry processing suite (Fig. 2).

At first sight the section is divided into three distinct intervals: (1) from the surface down to a fairly complex but strong reflector between 0.25 and 0.32 s TWT; (2) a relatively signal-free zone down to a strong band of reflections at 6 s TWT; and (3) down to the full recording time including many strong but impersistent phases. The complete record is generally laterally homogeneous except for a small region between 1.6 and 1.9 s TWT in the signal-free interval, which includes a few high frequency continuous reflections at the north end of the profile. These reflections are associated with anomalously high stacking velocities. Throughout the section, although most of the reflections are almost horizontal, some do show dips up to a maximum of ~100 ms in 2 km, giving maximum apparent dips of ~10°.

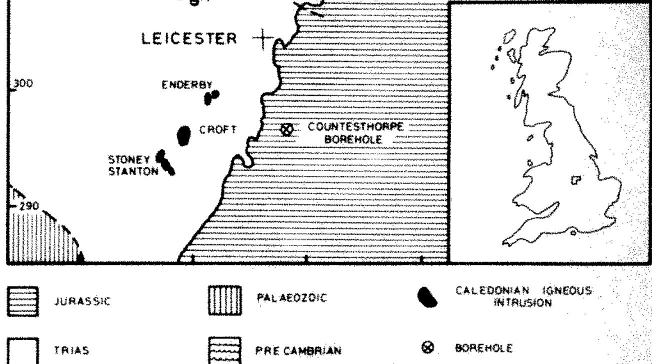


Fig. 1 Location map.

To examine the variation of reflection continuity with depth, a plot of length of continuous, correlatable signal (normalized to the maximum observed length averaged over a 0.2 s window) versus two-way time was produced (Fig. 3).

The anomalous high-frequency reflections between 1.6 and 1.9 s TWT were omitted being of very short lateral extent, and therefore uncharacteristic of the whole profile. The two minor peaks at ~0.6 and 0.9 s TWT are almost certainly multiples. The signal-free zone is divided into an upper, completely free zone down to 4 s TWT and a lower zone including a few, short, continuous correlatable pulses. Below the 6 s TWT band of reflections a second prominent group at 7.2 s TWT appears at the top of the broad zone of strong but impersistent reflections. To estimate the depth of the various reflection zones, a velocity–depth profile has been derived from other data.

The granite velocity, derived from Birch[4] is compatible with that observed near the upper surface of the Mountsorrel granodiorite[5]. Lower crustal and sub-Moho velocities, together with a reasonable depth to the Moho of 30 km, were obtained from the LISPB results[6], which are not inconsistent with a possible velocity structure beneath Charnwood Forest[7] (Fig. 1). Using these velocities it is necessary to invoke a change from 6.3 to 7.1 km s^{-1} at mid-crustal levels. It is considered that such a change could occur at the major reflection band at 6 s TWT. Although the velocity jump and associated two-way time are purely speculative, there is supporting evidence in the fact that interval velocities derived from the stacking velocities of the three most prominent reflections at 0.25, 6.0 and 7.2 s TWT are very similar to the constructed velocity–depth profile (Fig. 4). This has been used to generate the depth scale on Fig. 3. Below 7.2 s TWT, the identification of any coherent events as primary reflectors would be highly speculative, particularly in view of the very limited horizontal subsurface cover.

There are several more conclusions to be drawn from this study. The Leicestershire granite pluton does provide an acoustically transparent medium through which seismic energy can penetrate the lower crust. However, it is not possible to identify positively the base of granite. A slight change in character of the section below 4 s TWT may indicate a different crustal fabric but the main change occurs below 6 s TWT. The Moho, at a depth of ~30 km, almost certainly lies within the broad zone of deep reflections below the two prominent reflection bands at 6 and 7.2 s TWT. Although it is unreasonable to identify primary reflectors in this broad zone, and therefore unreasonable to speculate on the precise nature of the crustal–upper mantle transition, the presence of the deep reflections does imply that there are changes in acoustic impedance over a broad depth zone extending into the upper mantle, in contrast to the uniform structure apparent in the upper part of the section below the top surface of the granite.

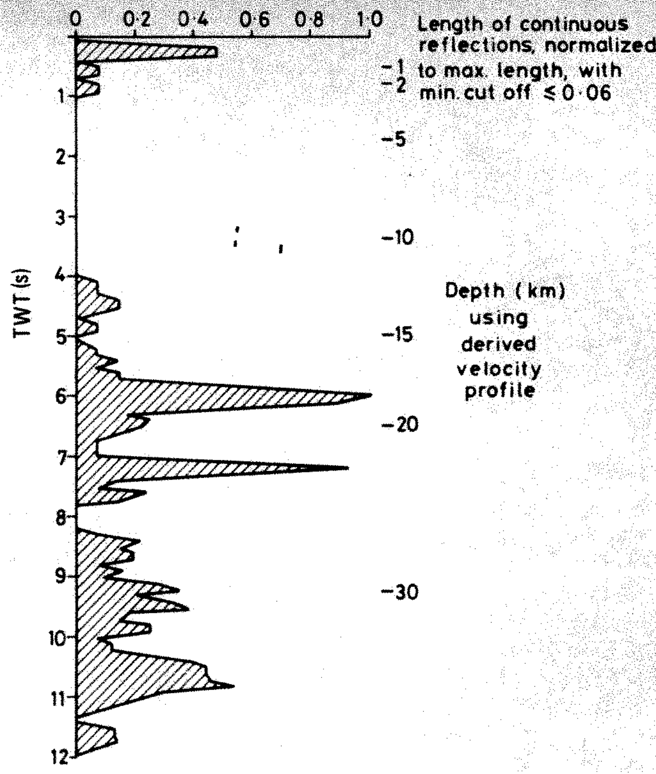


Fig. 3 Section reflection character.

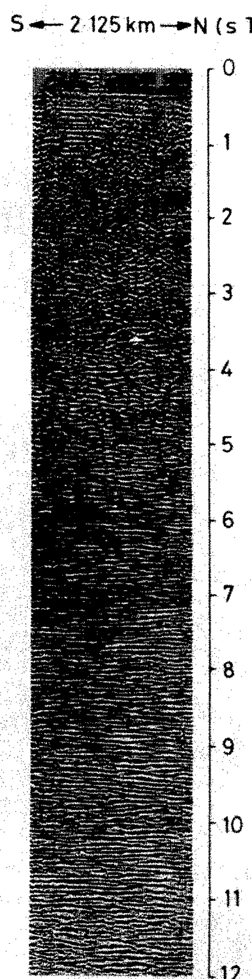


Fig. 2 Seismic section.

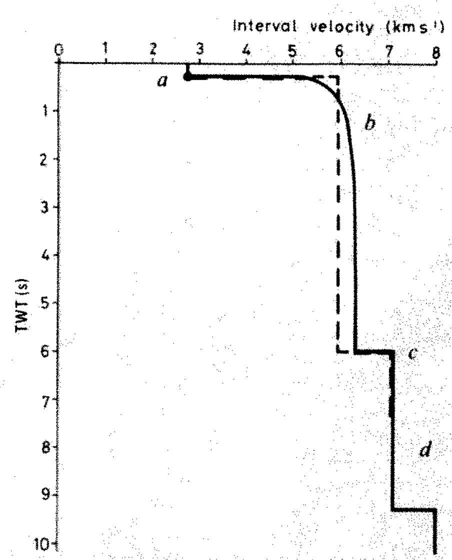


Fig. 4 Velocity–depth profile. – – –, Velocities from V_s for prominent reflections only. ———, Velocity profile used to convert from time to depth, derived from: *a*, borehole log; *b*, granite; *c*, break occuring at prominent reflector at 6.0 s TWT; *d*, LISPB and study of P-wave delay between Charnwood and Eskdalemuir.

This paper is published with the approval of the Director, Institute of Geological Sciences (NERC).

Received 22 December 1981; accepted 30 April 1982.

1. Pidgeon, R. T. & Aftalion, M. in *Crustal Evoluion in Northwestern Britain and Adjacent Regions* (eds Bowes, D. R. & Leake, B. E.) 183–220 (Seel, Liverpool, 1978).
2. Le Bas, M. J. *Trans. Leicester Lit. phil. Soc.* **75**, 41–57 (1981).
3. Maguire, P. K. H., Whitcombe, D. N. & Francis, D. J. *Trans. Leicester Lit. phil. Soc.* **75**, 58–66 (1981).
4. Birch, A. F. in (eds Benioff, V. H. *et al.*), *Contr. Geophys. Int. Ser. Mongr. Earth. Sci.* Vol 1, 158–170 (Pergamon, New York, 1958).
5. Whitcombe, D. N. & Maguire, P. K. H. *J. geol. Soc. Lond.* **138**, 643–651 (1981).
6. Bamford, D., Nunn, K., Prodehl, C. & Jacob, B. *Geophys. J. R. astr. Soc.* **59**, 43–60 (1978).
7. Maguire, P. K. H., Bermingham, P. M. & Francis, D. J. *Geophys. J. R. astr. Soc.* **65**, 229–235 (1981).

Methane flux in the Great Dismal Swamp

Robert C. Harriss & Daniel I. Sebacher

Mail Stop 270, NASA Langley Research Center, Hampton, Virginia 23665, USA

Frank P. Day Jr

Department of Biological Sciences, Old Dominion University, Norfolk, Virginia 23508, USA

Methane is an important component of the biogeochemical cycle of carbon with potentially critical roles in both atmospheric chemical and radiation transfer processes[1–4]. Limited evidence is available which suggests an increase in global tropospheric methane during the last decade[5,6]. To understand and assess the possibility and implications of temporal variations in atmospheric methane requires improved quantitative knowledge of methane sources and sinks. We report here methane flux measurements made over a 17-month period in the Great Dismal Swamp, Virginia. These flux measurements indicate that Great Dismal Swamp soils can act as both a source and sink for atmospheric methane. In a waterlogged condition, swamp soils are a net source of methane to the atmosphere with seasonal variations in emission rates from <0.001 to 0.02 g CH_4 m^{-2} day^{-1}. During drought conditions, swamp soils consume atmospheric methane at rates of <0.001 to 0.005 g CH_4 m^{-2} day^{-1}. While these results should not be extrapolated to all swamp soils, they illustrate the potential complexity of processes which regulate net flux of methane between wetland soils and the atmosphere.

Major natural sources of atmospheric methane involve anaerobic fermentation of organic material by microbial activity in wetland ecosystems and enteric fermentation in mammals[2]. While several laboratory and field investigations have identified variables correlated and coupled to the biogenesis of methane in soils and sediments[7–13], existing data are inadequate as a basis for quantitative estimation of global sources. Inadequate methodology for *in situ* quantitative measurements of methane flux at air–water and soil–water interfaces in wetland environments has been a primary limiting factor. Previous measurements have involved limited sampling of gas bubble flux through a water column or use of simple static chamber techniques. Flux measurements in static chambers have been criticized for producing serious modification of the atmosphere over the soil surface during long incubation periods. We have developed a new technique for gas flux measurements in wetland environments which combines a gas-filter-correlation (GFC) IR absorption analyser with improved sampling chambers that enclose a soil area (closed chambers) or chambers that are open to artificially induced air flow from outside (open chambers). This continuous sampling and analysis system has high precision (±0.01 p.p.m.) and fast response time (1 s) for measuring methane flux as low as 1×10^{-4} g CH_4 m^{-2} day^{-1} in <20 min. The system is portable and floatable and is operated at remote locations using battery power. Flux measurements reported here were obtained with a 0.5 m^3 closed chamber. In the closed chamber mode, air continuously recirculates from the sampling chamber over the soil through the GFC and back to the chamber; any changes in methane concentration in the chamber are measured continuously. A circulation fan maintains an average air flow of 0.63 m s^{-1} in the chamber. Chamber temperature and humidity are maintained at ambient atmospheric values throughout a flux measurement. Calibration of the system is carried out each day in the field using certified methane standards. Details of instrument design, calibration and field testing are given elsewhere[14].

A study of plant ecology, soils and nutrient cycling in the Great Dismal Swamp, Virginia (latitude 30°40′ N, longitude 76°30′ W) has provided an excellent opportunity to examine ecological and climatological factors in relation to temporal variations in methane flux. Extensive background data are available on the hydrology, ecology and land-use history of the Great Dismal Swamp[15–19]. Measurements of water levels and soil and air temperatures were collected throughout our 17-month study period.

The Great Dismal Swamp is an 85,000-hectare, forested, peat bog and is one of the more extensive swamp forests still remaining in the USA[20]. All of the methane flux measurements reported here were taken in a maple-gum habitat. Maple-gum (*Acer rubrum* and *Nyssa aquatica*) is the dominant vegetation in the swamp, with scattered stands of cypress (*Taxodium distichum*) and cedar (*Chamaecyparis thyides*) as secondary components. Seasonal inundation of the maple-gum habitat varies in depth and frequency reflecting local rainfall. During the period of this study, we had an opportunity to study the

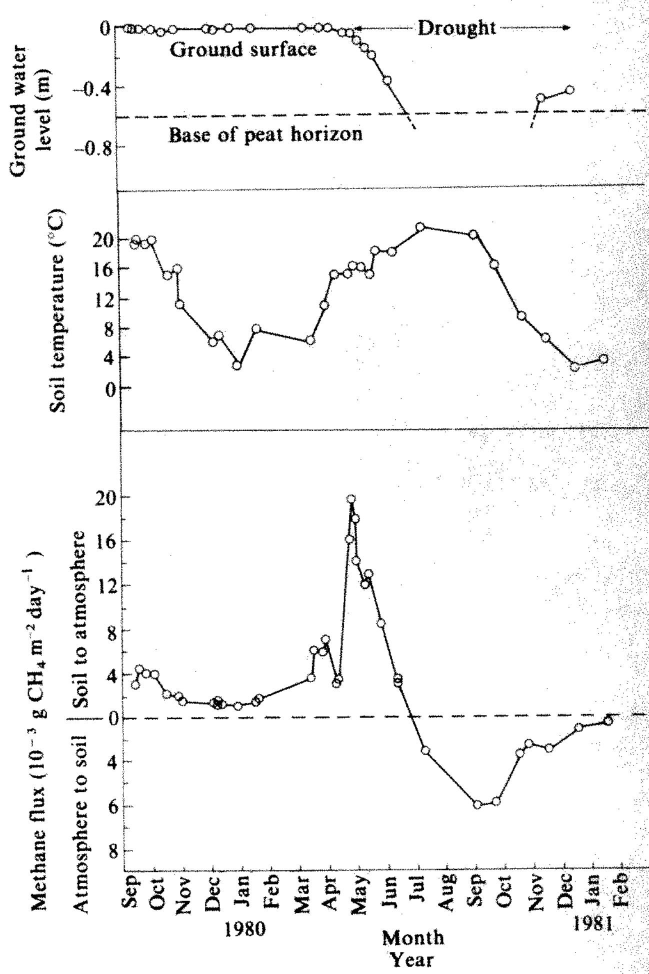


Fig. 1 Measured ground water levels, soil temperatures at 5 cm and methane flux in a Great Dismal Swamp peat soil between September 1979 and February 1981.

swamp in conditions ranging from totally flooded soils to dry soils resulting from drought conditions. All fluxes reported here are from soil or water surfaces without vegetation in the chamber.

Figure 1 summarizes temporal variations in water levels, soil temperature and methane flux for the Dismal Swamp maple-gum soil from September 1978 to February 1981. Two features of the methane flux variability are of particular interest. First, measured emissions of methane from waterlogged soils to the atmosphere ranged from a low of 1.3×10^{-3} g CH_4 m^{-2} day^{-1} in December to a high of 19.7×10^{-3} g CH_4 m^{-2} day^{-1} in April. From September to December, methane emission rate decreased as soil temperature decreased. During March and April, methane emission rate increases as soil temperature increases. Note that for any specific soil temperature above 2°C, methane emissions in spring exceed fall rates.

Drawing analogy with studies in lakes, we hypothesize that microbial activity in the soil profile is stratified with methane oxidizing microorganisms living in the shallow aerobic surface layer and methane producing microorganisms living at depth in oxygen depleted habitat. Our measurements are net flux between the soil and atmosphere. During winter surface soils down to ~2 cm are at temperatures <1°C, while the remainder of the peat profile (2–60 cm) maintains temperatures of 2–4°C. Thus, the early spring maximum in methane flux to the atmosphere in April 1980, may represent a period when winter–spring production relative to oxidation rates reaches a maximum. Snow cover and frozen soils would both inhibit metabolic activities of methane oxidizing microorganisms in surface soils and reduce gas exchange across the soil–air interface, leading to an accumulation of methane in the soil during winter months. The pattern of methane evolution we observed in spring 1981 in the Great Dismal Swamp may be similar in origin to episodic gas evolution from Arctic tundra soils during spring thaw[21].

Methane fluxes to the atmosphere from maple-gum peat soils in the Dismal Swamp are generally lower than flux rates measured for other wetland soils[13,22–24]. Several factors including the enhancement of emissions by aquatic plants[25], differences in soil organic matter and nutrient content[26], and differences in measurement techniques could be contributing to the observed differences in flux rates. We do not claim that our measured flux rates in the Dismal Swamp are representative of other forested wetlands; rather we consider the primary importance of our data is in illustrating the potential complexity of processes which regulate long-term net flux of methane between wetland ecosystems and the atmosphere.

The second major feature in our study of temporal variations in methane flux is the reversal of the flux from July 1980 to February 1981. During this period, the peat soils in the maple-gum habitat of the Great Dismal Swamp were a sink for atmospheric methane. Methane removal from ambient air over the soil surface is associated with drought conditions. The water table dropped below the bottom of the surface peat, producing a relatively dry, porous soil. From October 1980 to February 1981, the methane flux from the atmosphere to the soil decreased as soil temperature decreased. While methane oxidation in certain lake and marine environments is well documented[27–31], we believe this is the first field measurement of a soil sink for atmospheric methane. Previous investigators concerned with the atmospheric methane budget have concluded that the soil is not a significant sink[32,33]. While our results represent only one ecosystem, the question of dry peat soils as a sink for atmospheric methane should be considered an unresolved issue. Measurements over a normally dry soil under hardwood forest during this same time interval showed no detectable methane flux.

As an independent approach to the study of methane consumption in soil, we sampled gas from depths of 5, 12 and 33 cm in the peat soil. Small volume samples were drawn into evacuated flasks, followed by laboratory gas chromatographic analysis. Methane concentrations in soil gas decrease with depth from 1.6 p.p.m.v. (surface), 1.5 (5 cm), 1.4 (12 cm), to 1.0 p.p.m.v. (33 cm). Using these data to calculate methane flux from atmosphere to soil, assuming molecular diffusion is the primary transfer process, a value of 1.8×10^{-3} g CH_4 per m^2 per day is obtained. Our measured net flux at the soil–air interface using the GFC system on the same day was 2.6×10^{-3} g CH_4 per m^2 per day. The slightly higher value for the directly measured flux is expected because factors such as the air flow over the soil surface can enhance transfer rates relative to molecular diffusion. The relatively close agreement between independent techniques adds further support to the concept of Dismal Swamp peat soils as a sink for atmospheric methane during dry conditions.

Our results raise questions concerning the generally accepted estimates of global methane emissions from wetlands and the assumption that soils are not a significant sink for atmospheric methane. Previous estimates of methane flux from major wetlands of the world have been based on a simple temperature–methane production relationship[2,13]. Our results indicate that soil water content, temperature and other seasonal climatological factors are all potentially critical factors in determining whether a wetland soil is a net source or sink for atmospheric methane. Most of the major tropical freshwater wetlands of the world are subject to seasonal fluctuations in rainfall and water levels. The Florida Everglades, Okefenokee Swamp, Sudd and most rice fields are typically dry for at least several months of the year. In addition, extensive land reclamation projects are underway in the Everglades, Sudd and other wetlands which involve drainage and conversion of wetland soils to agriculture. These considerations indicate that a reassessment of the role of natural wetlands as a source of methane to the global atmosphere is necessary. Most importantly, our results illustrate the potential complexity of the problem of quantitatively determining long-term net flux of methane from a wetland ecosystem to the atmosphere.

We thank the staff of the Dismal Swamp National Wildlife Refuge for cooperation. Ecological studies in the swamp are supported by the NSF under grant DEB-77-08609 to F.P.D. Methane flux studies are supported by NASA, Life Sciences Division, Global Ecology Program under a grant to R. C. H. and D. I. S. Dr C. S. Martens provided detailed comments on the manuscript and suggested flux calculations using soil gas concentrations.

Received 16 March; accepted 27 April 1982.

1. Crutzen, P. J. in *Physics and Chemistry of Upper Atmospheres* (ed. McCormac, B. M.) 110 (Reidel, New York, 1973).
2. Ehhalt, D. H. & Schmidt, U. *Pure appl. Geophys.* **116,** 452–464 (1978).
3. Wofsy, S. C. *A. Rev. Earth planet. Sci.* **4,** 441–469 (1976).
4. Sze, N. D. *Science* **195,** 673–674 (1977).
5. Gradel, T. E. & McRae, J. E. *Geophys. Res. Lett.* **7,** 977–979 (1980).
6. Rasmussen, R. A. & Khalil, M. A. K. *J. geophys. Res.* **86,** 9826–9832 (1981).
7. Koyama, T., *J. geophys. Res.* **68,** 3971–3973 (1963).
8. Mah, R. A., Ward, D. M., Baresi, L. & Glass, T. L. *A. Rev. Microbiol.* **31,** 309–340 (1977).
9. Zeikus, J. G. *Bacter. Rev.* **41,** 514–541 (1977).
10. Atkinson, L. P. & Hall, J. R. *Estuar. Coast. mar. Sci.* **4,** 677–686 (1976).
11. Martens, C. S. & Berner, R. A. *Science* **185,** 1167–1169 (1974).
12. Swain, F. M. *Adv. org. Geochem.* **1,** 673–687 (1973).
13. Baker-Blocker, A., Donahue, T. M. & Mancy, K. H. *Tellus* **29,** 245–250 (1977).
14. Sebacher, D. I. & Harriss, R. C. *J. envir. Qual.* **11,** 34–37 (1982).
15. Whitehead, D. R. *Ecol. Monogr.* **42,** 301–315 (1972).
16. Carter, V., Garrett, M. K., Shima, L. & Gammon, P. *Wat. Res. Bull.* **13,** 1–12 (1977).
17. Day, F. P. & Dabel, C. V. *Vac. J. Sci.* **29,** 220–224 (1978).
18. Dabel, C. V. & Day, F. P. *Bull. Torrey bot. Club* **104,** 352–360 (1977).
19. Day, F. P. *Am. Mid. Nat.* **102,** 281–289 (1979).
20. Oosting, H. J. *The Study of Plant Communities* (Freeman, San Francisco, 1956).
21. Gosink, T. A. & Kelley, J. J. *J. geophys. Res.* **84,** 7041 (1979).
22. Harriss, R. C. & Sebacher, D. I. *Geophys. Res. Lett.* **8,** 1002–1004 (1981).
23. Cicerone, R. J. & Shetter, J. D. *J. geophys. Res.* **86,** 7203–7209 (1981).
24. Harriss, R. C., Sebacher, D. I., Bartlett, K. D. & Bartlett, D. S. *Proc. Sym. Comp. Nonurban Troposphere*, (American Meteorological Society, in the press).
25. Dacey, J. W. H. & Klug, M. J. *Science* **203,** 1253–1255 (1979).
26. Kelly, C. A. & Chynoweth, D. P. *Limnol. Oceanogr.* **26,** 891–897 (1981).
27. Rudd, J. W. M. & Hamilton, R. D. *Limnol. Oceanogr.* **23,** 337–348 (1978).
28. Sansone, F. J. & Martens, C. S. *Limnol. Oceanogr.* **23,** 349–355 (1978).
29. Barber, L. E. & Ensign, J. C. *Geomicrobial J.* **1,** 341–345 (1979).
30. Fallon, R. D., Harrits, S., Hanson, R. S. & Brock, T. D. *Limnol. Oceanogr.* **25,** 357–360 (1980).
31. Harrits, S. M. & Hanson, R. S. *Limnol. Oceanogr.* **25,** 412–421 (1980).
32. Whittenbury, R. H., Dalton, R. H., Eccleston, M. & Reed, H. L. in *Proc. Congr. on Microbial Growth in C_1 Compounds* 1–9 (1974).
33. Ehhalt, D. H. & Heidt, L. E. *J. geophys. Res.* **18,** 5265–5271 (1973).

Evidence of colonial nesting and 'site fidelity' among ornithischian dinosaurs

John R. Horner

Museum of the Rockies, Montana State University, Bozeman, Montana 59715, USA

Recent discoveries in the late Cretaceous (Campanian) Two Medicine Formation of western Montana indicate that some dinosaur species, like some modern species of birds and crocodiles, nested in colonies. I report here evidence that members of one species returned to the same nesting area for many years. There are also indications that, after hatching, some species remained in their respective nests whereas others left the nest but remained in, or at least returned occasionally to, the nesting site. The area in which these discoveries were made is the Willow Creek Anticline of Teton County, where in 1978 a nest of fifteen, 1-m long hadrosaurs provided the first evidence of extended parental care among dinosaurs[1].

I have found in the same sediment horizon that produced the '1978 nest', a second nest of younger juvenile hadrosaurs (0.5 m long) and the weathered remnants of six unoccupied nests containing abundant eggshell remains (Fig. 1). All of the eight nests seem to have hosted conspecific hadrosaurs, of an as yet undetermined species, as the eggshell fragments from each site are structurally identical. Each nest appears to have been constructed as a circular or oval pit within a preconstructed mud mound (see ref. 1; Fig. 1). They were each found at an equal depth beneath a palaeosol, incorporated in a grey-brown mudstone interpreted as a floodplain deposit[2]. Where erosion or cover has not destroyed or obscured parts of the nesting horizon, the nests are at least 7 m apart or approximately a distance equivalent to the length of an average adult hadrosaur from this area. There may, therefore, have been as many as 40 nests in the 10,000 m^2 area. The occurrence of eight nests along what appears to have been a single time horizon, within a relatively small area, suggests that these hadrosaurs were nesting in a colony.

With the exception of a single isolated juvenile in the nesting area, the remains of the young hadrosaurs were found in nests associated with eggshell. Although the skeletal elements of the young are relatively well ossified, even on the youngest individuals, there is no indication that they were precocial self-feeding individuals that would have ventured out of their respective nests without parental supervision, as suggested for

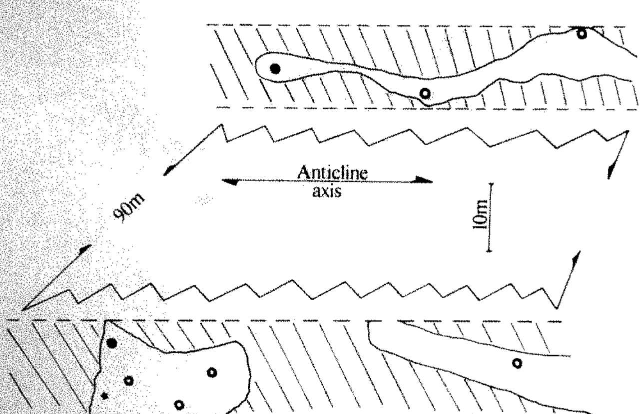


Fig. 1 Relationship of hadrosaur nests straddling Willow Creek Anticline. ●, Nests containing juveniles; ○, nests with eggshell concentrations and the asterisk indicates the location of the isolated juvenile. Broken lines represent the limit of the nest-bearing horizon. Cross-hatched areas are covered or missing.

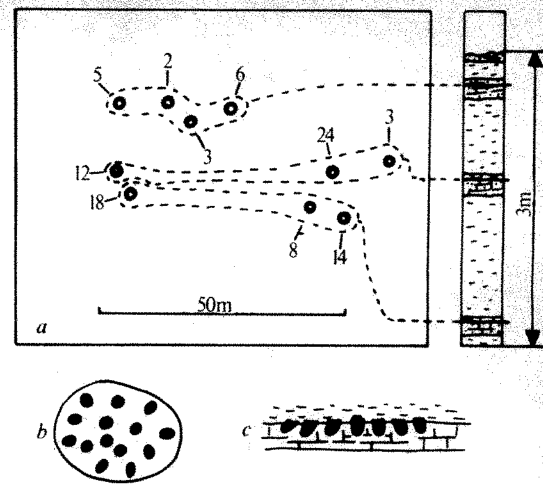


Fig. 2 *a*, Map and vertical section of the anticline showing the relationship of the egg clutches attributed to the hypsilophodon-like ornithopod. Values represent the number of eggs per nest, broken lines enclose clutches found on single horizons. *b*, Typical clutch arrangment viewed from above. *c*, Egg clutch viewed from the side showing how the lower portions of the eggs are located in siliceous carbonate sediment.

Psittacosaurs[3]. Early ossification may instead simply reflect rapid growth. As for the juveniles being too small for effective feeding by the adults, or possibly being crushed by an adult 'mis-step' (ref. 3), the fact that the crocodilians manage to overcome these difficulties[4–6] implies that other archosaurs could also manage to do so. If by instinct or parental assistance, the young remained in their respective nests and the parent fed them from the nest perimeter, the young may have grown more rapidly, and there would have been less risk of infant mortality. Protection against predators would have been afforded by both nest confinement of the young and the closely-packed nests of the colony with the continued presence of many adults. As some adults foraged and gathered food for their young, the presence of other adults remaining in the nesting area may have thwarted potential predators, as is thought to be the case for many colonizing birds[7]. The presence of the two nests containing the remains of juveniles probably reflects this security in that they were not obviously preyed upon. Their deaths may have been due to starvation after the death of their parents.

A second nesting site (Fig. 2) on Willow Creek Anticline which has yielded several juvenile ornithopods, closely allied to the Hypsilophodontidae, offers evidence not only of colonial nesting but also of 'site fidelity' (multi-year use of nesting site) and the possibility of bird-like crèches. The 10 nests attributed to these ornithopods contain the hatched remains of up to 24 eggs per clutch. Clutches that do not appear to have suffered erosional damage are ~1 m in diameter, and were found along at least three different horizons, incorporated in brown, organic-rich, siliceous carbonates with mudstone inclusions[2]. These are interpreted as products of soil formation where partially dried sediments were repeatedly disrupted and mixed in the process of nest construction and daily traffic[2]. Although the exposed site of these nests is being excavated, the number of clutches so far found along these three horizons suggests colonial nesting for this species. The fact that they occur at different levels indicates that the members of the species returned to the site for many years. The elongate, ellipsoid eggs attributed to these ornithopods were meticulously laid with their smaller ends 'planted' upright to oblique, partially into the siliceous carbonate sediments (Fig. 2). On hatching, the young left through the top portion of the eggs, leaving the lower portions intact within the sediment. The fact that the lower portions were preserved rather than crushed and broken by trampling of the young, together with the fact that juvenile remains were not found in the nests, suggests that the young of this species did not remain in the nests long after hatching. It is possible, however, that

0028-0836/82/250675—02$01.00

the juveniles from different nests assembled into bird-like crèches, as the remains of at least 12 small skeletons of varying sizes, have been found on the nesting horizons. Adults of this species appear to have been ~2.5–3 m long. The juvenile remains represent individuals ~0.5–1.5 m long, indicating that the young of this species either remained in the colony or returned to the site frequently, either of which may have been a result of parental care.

Considering the variety of nesting habits and social behaviours exhibited by the different crocodilian species or particularly by Recent birds, it is expected that the morphologically and ecologically diverse dinosaurs would also have exhibited a variety of behaviours.

I thank the 1979, 1980 and 1981 Princeton Expedition members for their hard work and many discoveries, D. Schankler and D. Wilcove for their helpful comments and the NSF (grant EAR 8007817) for support.

Received 22 March; accepted 5 May 1982.

1. Horner, J. R. & Makela, R. *Nature* **282**, 296–298 (1979).
2. Lorenz, J. C. thesis, Princeton Univ. (1981).
3. Coombs, W. P *Nature* **283**, 380–381 (1980).
4. Hunt, R. C. *Copeia* **1975**, 763–764 (1975).
5. Pooley, A. C. *Animal Kingdom* **79**, 7–13 (1976).
6. Pooley, A. C. & Gans, C. *Scient. Am.* **234**, 114–124 (1976).
7. Welty, J. C. *The Life of Birds*, 546 (Knopf, New York, 1963).

Lucy's limbs: skeletal allometry and locomotion in *Australopithecus afarensis*

William L. Jungers

Department of Anatomical Sciences, School of Medicine, State University of New York at Stony Brook, Long Island, New York 11794, USA

Precise information about the bodily proportions of early hominids is crucial for accurate functional and phylogenetic interpretations of early human evolution[1–6]. The partial skeleton of *Australopithecus afarensis* (AL 288-1; 'Lucy'[7,8]) recovered in 1974 from the Hadar area of Ethiopia[9] permits the first direct assessment of body size, limb proportions and skeletal allometry in ancestral hominids that pre-date 3 Myr. Using allometric relationships for limb lengths in non-human catarrhine primates (as a whole and for African apes alone) as empirical base lines for comparison, I show here that the limb proportions of *A. afarensis* are clearly unique among hominoids. The data indicate that *A. afarensis* had already attained forelimb proportions similar to those of modern humans but possessed hindlimbs that were relatively much shorter; hence the 'intermediate' humerofemoral index of AL 288-1 (85.1) compared with *Homo sapiens* and great apes[9,10]. It follows that relative and absolute elongation of the hindlimbs represents one of the major evolutionary changes in later human evolution. The bodily proportions of Lucy are not incompatible with some form of bipedal locomotion, but kinematic identity and functional equivalence with the bipedal gait of modern humans seem highly improbable. Reduced relative stride length in AL 288-1 probably implies both greater relative energy cost and relatively lower peak velocities of bipedal locomotion in *A. afarensis*.

Speculation about the limb proportions of the earliest hominids has long preoccupied palaeoanthropologists and functional anatomists[1–6,11–15], largely due to the obvious importance of such information for realistic reconstructions of locomotor behaviour and probable phylogenetic pathways. Due to the lack of associated, complete limb bones from both fore- and hindlimbs of *Australopithecus*, the earliest unequivocal hominid, comparisons of bodily proportions with modern *H. sapiens* have been necessarily tentative and inconclusive. Some authors suspect that australopithecines possessed proportions virtually identical to those of modern humans, perhaps for the entire postcranial skeleton[16,17] or just for the hindlimb[2,18]. Others have proposed that the forelimbs were relatively long[7,12,15] and/or the hindlimbs relatively short[4,6,12,14]. The recently prepared, partial skeleton of *A. afarensis* (AL 288-1; 'Lucy') includes complete elements of forelimb (humerus) and hindlimb (femur), and therefore provides the first opportunity to test previous inferences about limb proportions and skeletal allometry in ancient hominids. As Walker[3] notes, it is only when such data are available that "biomechanical analyses of the postcranial parts of *Australopithecus* will become more convincing than they are at present".

Based on a skeletal sample of 454 adult, non-human catarrhine primates (Old World monkeys, lesser apes and great apes) and wildshot body weights, I have established allometric relationships for humerus length and femur length in catarrhines as a whole (a size spread from *Miopithecus talapoin* to *Gorilla gorilla*) and in African apes alone (from female *Pan paniscus* to male *G. gorilla*)[19]. The sample includes 164 cercopithecines, 77 colobines, 116 hylobatid apes and 97 pongids. Body weights are taken from museum records and the literature. These scale relationships serve in this analysis as comparative, empirical baselines from which vertical deviations (d_{yx}) can be identified in AL 288-1, a sample of modern *H. sapiens* and a small female bonobo chimpanzee. For prediction and d_{yx} calculation, model I regression of $\log_{10}$-transformed data is regarded as more appropriate[19–21] than model II methods (although principal axis solutions yield identical conclusions). Least-squares regression of $\log_{10}$ (humerus length) and log (femur length) on $\log_{10}$ (body weight)$^{1/3}$ yields the following predictions for catarrhines as a whole:

$$\text{Humerus:}\quad \log_{10}Y = 1.048\log_{10}X + 0.862\ (r = 0.88)$$
$$\text{Femur:}\quad \log_{10}Y = 0.828\log_{10}X + 1.192\ (r = 0.94)$$

For African apes only, the predictive equations are:

$$\text{Humerus:}\quad \log_{10}Y = 0.815\log_{10}X + 1.229\ (r = 0.99)$$
$$\text{Femur:}\quad \log_{10}Y = 0.437\log_{10}X + 1.806\ (r = 0.94)$$

Following Smith[17,21], percentage deviations ('prediction errors') for humerus length and femur length were computed from each equation as

$$\frac{\text{Observed value} - \text{predicted value}}{\text{Predicted value}} \times 100$$

for AL 288-1, a female Mbuti pygmy *H. sapiens* (MMC-18)[22], male and female samples of caucasian *H. sapiens* and a female bonobo *P. paniscus* (T-29060). Body weight of Lucy has been estimated by her discoverers at 60 lb (27.3 kgf)[7]. A conservative 25.0–30.0 kgf range has been used in this analysis, consistent with independent estimates derived from both linear[23] and multiple (C. O. Lovejoy, personal communication) regression. Body weight of the Mbuti pygmy was calculated elsewhere[24], and the caucasian male and female values represent actual weights at death (Table 1). At 27 kgf the wildshot female bonobo falls at the lower limit of adult body size in *P. paniscus*. It is included in this analysis specifically because it is a pongid skeleton that is similar in size to both AL 288-1 and MMC-18.

Percentage deviations (Table 1) disclose a close similarity between AL 288-1 and both *H. sapiens* groups in relative humerus length. The humerus is slightly shorter than predicted in all the hominids, whereas that of the female bonobo is longer than expected. In contrast, prediction errors for femur length indicate a great disparity between modern humans on the one hand and Lucy and T-29060 on the other. With the exception of AL 288-1 relative to African apes, femoral deviations are positive in all cases, but those of the modern human groups are much greater than Lucy's. The same results obtain even if an estimate of 50 lb (22.7 kgf) is used for Lucy's weight. Intergroup allometric projections (Table 2) reinforce these observations. An AL 288-1 to caucasian projection results in near isometry ($k = 1.10$) for humerus length but extreme positive

0028–0836/82/250676—03$01.00

Table 1 Deviations (d_{yx}) from allometric relationships in extant catarrhine primates

	BW (kgf)	HL (mm)	FL (mm)	% Deviation from catarrhine line HL	% Deviation from catarrhine line FL	% Deviation from African ape line HL	% Deviation from African ape line FL
A. afarensis (AL 288-1) ♀	25.0–30.0	239	281	−4.5 to −10.4	+4.8 to +10.2	−9.9 to −14.3	−2.1 to +0.1
H. sapiens Mbuti pygmy ♀ (MMC-18)	27.3–30.9	256	349	−0.8 to −5.0	+29.1 to +33.6	−5.8 to −8.9	+21.1 to +23.3
H. sapiens							
Caucasian ♂ ($n = 9$)	68.3 (±5.2)	336 (±17)	469 (±27)	−5.6	+39.4	−3.6	+44.9
Caucasian ♀ ($n = 9$)	55.0 (±5.5)	308 (±11)	433 (±18)	−6.6	+36.6	−6.3	+38.1
P. paniscus (T-29060) ♀	27.0	289	291	+12.5	+11.6	+6.7	+2.9

Percentage deviation calculated as (observed value − predicted value)/(predicted value) × 100 (see refs 17, 21). Catarrhine allometric equations are model I regression estimates for a size run from *M. talapoin* to *G. gorilla*; African ape allometry equations are model I regression estimates for a size run from female *P. paniscus* to male *G. gorilla*. Model II regressions (such as principal axis) yield similar conclusions, but are statistically inappropriate for calculations of vertical residuals (d_{yx}). The body weight of AL 288-1 has been estimated at 60 lb (27.3 kgf)[7]; a conservative range of 25–30 kgf is used here. The body weight values for MMC-18 are from ref. 17. Body weights of the caucasian skeletal sample are from the autopsy records of the Cleveland Museum of Natural History; only non-pathological (neither obese nor malnourished) individuals were measured. The body weight of T-29060 is taken from the field notes at the Musée Royal de l'Afrique Centrale in Tervuren. Lengths of humerus and femur of AL 288-1 are taken from reconstructions made by C. O. Lovejoy at the Cleveland Museum of Natural History. BW, body weight; HL, humerus length; FL, femur length.

allometry ($k = 1.72$) for femur length. A pygmy to caucasian projection, by contrast, reveals near isometry for both skeletal elements. Bonobo to caucasian scaling requires extreme relative reduction of the forelimb ($k = 0.43$) coupled with extreme positive allometry of the femur ($k = 1.58$), comparable with the Lucy–human vector. Because AL 288-1, MMC-18 and T-29060 had approximately equal body weight, a direct comparison of their limb dimensions produces the same conclusion in a slightly different manner. Lucy possesses the shortest humerus of the three, but the Mbuti pygmy humerus is only 7% longer; that of the bonobo is 21% longer. Lucy also possesses the shortest femur, but that of the bonobo is <4% longer; the femur of MMC-18 is almost 25% longer than Lucy's. As the crural indices (tibia/femur × 100) of African apes and humans are very similar[19,25], a short femur necessarily implies that the entire hindlimb was comparably short. All analyses lead to the same conclusion—that forelimb proportions of *A. afarensis* (as represented by AL 288-1) were already comparable with those of modern humans, but that the hindlimb of this early hominid was relatively short and proportionately similar to that of a small pongid. Earlier suspicions[4,6,12,14] of relatively short hindlimbs in *Australopithecus* are confirmed by these findings, but predictions[4,7,12,15] of relatively long forelimbs are not substantiated (at least in these gracile australopithecines). These results also account for the intermediate humerofemoral index (85.1) of AL 288-1 compared with *H. sapiens* and pongids[9,10].

Dramatic hindlimb elongation (absolutely and relatively) emerges as one of the major evolutionary changes from *A. afarensis* to modern humans. This finding contrasts with the recent assertion[23] that natural selection would be unlikely to favour relative hindlimb lengthening in a short-limbed ancestral hominid. A major proportional difference between Lucy and a comparably sized ape is the shorter forelimbs of the former. Analogy to the locomotor behaviour of great apes and basic biomechanics of climbing[19,26] suggest that the climbing capability of *A. afarensis* was probably diminished by such interlimb proportions. Unfortunately, we do not know whether Lucy's forelimb proportions are primitive or derived for African hominoids. Nevertheless, we do know that to derive *A. afarensis* from a relatively small-bodied hominoid with limb proportions similar to *P. paniscus* would entail significant forelimb reduction but little change in hindlimb dimensions[6,14]. Because we do not know whether the immediate ancestor of *A. afarensis* had proportions comparable with bonobos, this evolutionary scenario cannot be tested. Given Lucy's pongid-like hindlimb proportions, reconstructions of stature in *A. afarensis* derived solely from metric trends evinced by modern humans[27] also become less convincing.

Diagnostic details of the knee joint and bony pelvis of *A. afarensis* are compelling indicators of a bipedal adaptation[7,28–30], and fossilized footprints[27,31] provide dramatic evidence of bipedal locomotion in *A. afarensis*. However, the relatively short hindlimb of Lucy implies substantial kinematic differences in bipedal gait from the modern condition. Both relative and absolute stride length were necessarily reduced in AL 288-1. The range of flexion and extension at the hip would have had to have been much greater in Lucy to match the stride of, say, MMC-18, but nothing in the anatomy of the hip joint would indicate an exaggerated degree of potential excursion[32]. These findings probably also account for the relatively short stride reconstructed for the makers of the Laetoli footprints[33] without having to resort to hypothetical 'strolling' behaviour.

Relative elongation of the hindlimb in later hominids would permit increasing velocity of gait at only a slight increase in energy cost because increased speed can be achieved by increased stride length rather than by increases in cadence (step frequency)[34–37]. Selection to reduce the energy cost of bipedal locomotion[38] would therefore favour hindlimb elongation, although relatively long extremities are probably also advantageous for enhanced performance (such as greater maximum terrestrial velocity). Thus the hindlimb proportions of *A. afarensis* and other 'primitive' details of the foot, ankle and hip[32,39] cannot be construed as evidence of a fully modern adaptation to bipedalism[31,33,40].

I thank D. C. Johanson and C. O. Lovejoy for access to the human and Hadar collections in Cleveland and the numerous curators of primate skeletal collections in the USA and Europe, and Jack Stern, John Fleagle and Randall Susman for suggestions and criticisms. This work was supported in part by a SUNY University Award and by NSF grant BNS 79-24070.

Table 2 Intergroup allometry coefficients ($y = \beta x^k$)

Pairwise groups	Exponent Humerus	Exponent Femur
A. afarensis → Caucasian (AL 288-1) (means)	1.10	1.72
Mbuti pygmy → Caucasian (MMC-18) (means)	0.93	1.04
P. paniscus → Caucasian (T-29060) (means)	0.43	1.58

'Exponent' is estimated by the least-squares (model I) regression of $\log_{10}$ (limb length) on $\log_{10}$ (body weight)$^{1/3}$. The weight of AL 288-1 used here is 60 lb (27.3 kgf); that of MMC-18 is the midpoint value of 29.1 kgf. An exponential value of 1.0 indicates isometry or geometric similarity.

Received 12 November 1981; accepted 14 April 1982.

1. Campbell, B. G. *Human Evolution* (Aldine-Atherton, Chicago, 1966).
2. Robinson, J. T. *Early Hominid Posture and Locomotion* (University of Chicago Press, 1972).
3. Walker, A. C. *J. hum. Evol.* **2**, 545–555 (1973).
4. McHenry, H. M. *Am. J. phys. Anthrop.* **49**, 15–22 (1978).
5. Oxnard, C. E. *Uniqueness and Diversity in Human Evolution* (University of Chicago Press, 1975).
6. Zihlman, A. & Brunker, L. *Yb. phys. Anthrop.* **22**, 132–162 (1979).
7. Johanson, D. C. & Edey, M. A. *Lucy, The Beginnings of Human Kind* (Simon & Schuster, New York, 1981).
8. Johanson, D. C. & White, T. D. *Science* **203**, 321–330 (1979).
9. Johanson, D. C. & Taieb, M. *Nature* **260**, 293–297 (1976).
10. *Nature* **260**, 389 (1976).
11. Lovejoy, C. O. & Heiple, K. G. *Am. J. phys. Anthrop.* **32**, 33–40 (1970).
12. Leakey, R. E. F. *Nature* **242**, 170–173 (1973).
13. McHenry, H. M. *Science* **190**, 424–431 (1975).
14. Zihlman, A. L. *S. Afr. J. Sci.* **75**, 165–168 (1979).
15. Kay, R. F. *Science* **182**, 396 (1973).
16. Wolpoff, M. H. & Lovejoy, C. O. *J. hum. Evol.* **4**, 275–276 (1975).
17. Smith, R. J. *J. theor. Biol.* **87**, 91–111 (1980).
18. Szalay, F. S. & Delson, E. *Evolutionary History of the Primates* (Academic, New York, 1979).
19. Jungers, W. L. *The Lesser Apes: Evolutionary and Behavioural Biology* (eds Preuschoft, H., Chivers, D., Brockelman, W. & Creel, N.) (Edinburgh University Press, in the press).
20. Goldstein, S., Post, D. & Melnick, D. *Am. J. phys. Anthrop.* **49**, 517–532 (1978).
21. Smith, R. J. *J. hum. Evol.* **10**, 165–173 (1981).
22. Toerien, M. J. *S. Afr. J. med. Sci.* **19**, 97–104 (1954).
23. Aiello, L. C. *Aspects of Human Evolution* (ed. Stringer, C. B.) (Taylor & Francis, London, 1981).
24. Reed, C. A. & Falk, D. *Fieldiana, Geol.* **33**, 423–440 (1977).
25. Schultz, A. H. *Hum. Biol.* **2**, 303–438 (1930).
26. Jungers, W. L. *Am. J. phys. Anthrop.* **49**, 303–314 (1978).
27. White, T. D. *Science* **208**, 175–176 (1980).
28. Johanson, D. C., Lovejoy, C. O., Burstein, A. A. & Heiple, K. G. *Am. J. phys. Anthrop.* **44**, 188 (1976).
29. Lovejoy, C. O. *Am. J. phys. Anthrop.* **50**, 460 (1979).
30. Lovejoy, C. O. *Am. J. phys. Anthrop.* **52**, 250 (1980).
31. Leakey, M. D. & Hay, R. L. *Nature* **278**, 317–323 (1979).
32. Stern, J. T. Jr & Susman, R. L. *Am. J. phys. Anthrop.* (in the press).
33. Charteris, J., Wall, J. C. & Nottrodt, J. W. *Nature* **290**, 496–498 (1981).
34. Saunders, J. B., Inman, V. T. & Eberhart, H. D. *J. Bone Jt Surg.* **35A**, 543–558 (1953).
35. Taylor, C. R., Heglund, N. C., McMahon, T. A. & Looney, T. R. *J. exp. Biol.* **86**, 9–18 (1980).
36. Grieve, D. W. & Gear, R. J. *Ergonomics* **5**, 379–399 (1966).
37. Beck, R. J., Andriacchi, T. P., Kuo, K. N., Fermier, R. W. & Galante, J. O. *J. Bone Jt Surg.* **63A**, 1452–1457 (1981).
38. Rodman, P. S. & McHenry, H. M. *Am. J. phys. Anthrop.* **52**, 103–106 (1980).
39. Tuttle, R. H. *The Emergence of Man* (eds Young, J. Z., Jope, E. M. & Oakley, K. P.) (Royal Society & British Academy, London, 1981).
40. Latimer, B. M., White, T. D., Kimbel, W. H., Johanson, D. C. & Lovejoy, C. O. *J. hum. Evol.* **10**, 475–488 (1981).

Unchanged thermoregulatory set-point in the obese mouse

Harry J. Carlisle

Department of Psychology, University of California, Santa Barbara, California 93106, USA

Paul U. Dubuc

Sansum Medical Research Foundation, Santa Barbara, California 93102, USA

The genetically obese (*ob/ob*) mouse has been extensively studied as a model of early-onset obesity. These mice are characterized by a wide variety of endocrine and metabolic defects including hyperglycaemia, hyperinsulinaemia with insulin resistance, hyperlipidaemia and hyperphagia[1,2]. The marked obesity of this mutant cannot be attributed to overeating because it remains fat even when fed the same amount of food as its lean littermate[3,4]. The earliest indicators of the *ob/ob* genotype, low metabolic rate and low body temperature, are detectable 7–10 days after birth[5,6]. Hypothermia might, therefore, be a prime contributor to the development of obesity because a greater percentage of the diet will be diverted to fat rather than 'wasted' for thermogenesis[7–9]. As the obese mouse exhibits hypothermia over a wide range of ambient temperatures, it has been suggested that their set-point for body temperature regulation is lower than that of the lean mouse[10]. If this is the case the obese mouse should maintain a subnormal body temperature when required to work for exogenous heat in a cold environment. We now report that the set-point is unchanged in the obese mouse because it behaviourally defends a body temperature equivalent to that of the lean mouse during cold exposure.

Four pairs each of obese male and female and lean male and female C57BL/6J mice, 12–16 weeks old at the beginning of the experiments, were housed from birth in a colony room at 23 °C. The test apparatus consisted of a $24\times10\times18$ cm cage with a 3×5 cm Plexiglas lever mounted at one end. The naive mice were trained to press the lever until stable rates of heat intake were observed when the cage was placed in a freezer at -8 ± 2 °C. Each response activated two 250-W IR lamps mounted outside the cage and focused on the area around the lever. The total power dissipated by the lamps was controlled at 300 W, giving a radiant flux density of 180 mW cm^{-2} at the lever. The lamps remained on as long as the lever was depressed. The time that the heat lamps were on and the total number of responses were recorded on counters in an adjoining room. Body temperature was measured before and after each test session using a Yellow Springs telethermometer and 402 rectal probe (2.5 cm insertion). After training and one additional practice session, each mouse was tested for 2 h every 3 or 4 days until four tests were completed.

Obese mice placed in the −8 °C test environment did not maintain their body temperature at the initial value of 34.8 °C (Table 1). Instead, they raised their temperature by almost 2 °C to a level comparable with that of the lean littermates. Although normothermic, their rates of heat intake were lower than those of lean mice. Lean females (25.1 g) obtained the most heat (26.1 s of heat per min), while obese males (50.7 g) and females (48.4 g) obtained the least. These relationships between body weight and heat intake suggest that the size of the animal significantly affects its exogenous heat requirement. This supposition is supported by the very high correlation coefficient ($r=-0.93$) between body weight and heat intake. The higher rates of heat intake in the lean mice might also reflect a greater heat loss between responses because they were far more active than the obese mice. Invariably they sat on the lever in order to activate it and then explored the cage until the next response. In contrast, the obese pressed with their forepaws but remained huddled in front of the lever when the lamps were off. This behaviour would minimize heat loss and also maximize heat intake because some radiant energy is emitted from the filaments of the lamps as they cool. The obese mice also pressed very briefly as if testing the lever during the interval between longer responses, and thereby they emitted twice as many lever presses as the lean mice (Table 1).

In another 2 h session we determined the latency to normothermia in obese and lean mice. Rectal temperatures were measured 5, 15, 30, 60, 90 and 120 min after the start of a test period. Both lean and obese mice had substantially elevated body temperatures within 5 min of the beginning of the session (Fig. 1). Peak temperatures were seen 10 min later and the values remained near 37 °C for the last hour of the test. When the intensity of the heat lamps was varied, all animals responded by increasing heat intake when the wattage was decreased and by reducing it when the wattage was increased. Thus, both

Table 1 Pre- and post-test body temperatures and heat intake during tests at −8 °C

Group	n	Body temperature (°C) Before	Body temperature (°C) After	Heat intake: Responses min^{-1}	Heat intake: Heat min^{-1}
Lean	8	36.5 ± 0.17	36.9 ± 0.15	2.9 ± 0.46	24.8 ± 0.57
Obese	8	34.8 ± 0.19*	36.6 ± 0.15	6.1 ± 0.64*	20.6 ± 0.33*

Values are the mean ±s.e.m.
* Significant ($P<0.01$) difference by t-test between lean and obese groups.

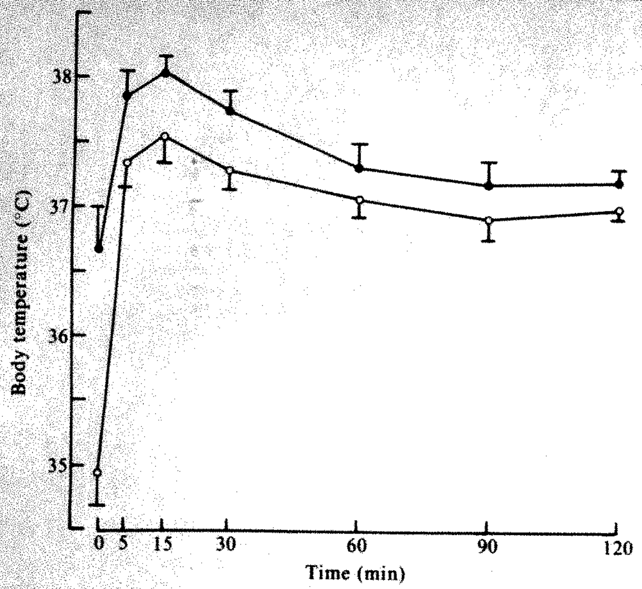


Fig. 1 Mean body temperatures (±s.e.m.) of eight lean (●) and eight obese (○) mice during the behavioural heat intake tests at −8 °C.

phenotypes responded appropriately and equally to variations in reward magnitude.

We also attempted to determine whether the trained mice would work for heat in a neutral ambient temperature. The mice were placed in the defrosted freezer in standard test conditions, except that the initial ambient temperature was 23±1 °C and could be raised if the mice worked for heat. In separate tests of 1 and 2 h, all mice raised the ambient temperature by working for heat. The obese mice raised the chamber temperature to 26.5 °C during the 1 h test and to 27.8 °C during the 2 h test. Their average body temperature at the end of the sessions was 36.0 °C, or 1 °C higher than their initial temperature. The lean mice raised ambient temperature to 29.0 °C during the 1 h test and to 30.9 °C during the 2 h test. Average body temperature increased from 36.9 to 37.4 °C. The lean mice obtained an average of 2.8 s of heat per min during the 1 h test and 2.2 s min^{-1} during the 2 h test, whereas obese mice obtained 1.2 and 0.84 s min^{-1} respectively. Thus, the obese mice raised their body temperature more but did not attain the same body temperature as the lean mice, nor did they obtain as much heat, or raise ambient temperature to the same level as the lean mice. We therefore conclude that the preferred ambient temperature of obese mice is 2.5–3.0 °C lower than that of lean mice. This effect may be due to the difficulty of increasing heat dissipation in a warm environment because of adiposity. It is also possible that the obese mouse, known for its low activity[11], may be unwilling to expend the effort necessary to raise ambient temperature above a minimally acceptable level. Other experiments are necessary to verify these alternatives.

The factors responsible for the hypothermia-hypometabolism of *ob/ob* mice remain to be defined. Our data indicate that hypothermia is not the result of a diminished requirement for heat or of an impaired sensitivity to ambient temperatures because the obese mouse behaviourally defends the same body temperature in a cold environment as the lean mouse. The diversion of dietary energy to fat rather than to heat must involve a defective or inappropriate heat production mechanism, possibly related to abnormal thyroid hormone production[12–15], or depressed hepatic[16] or brown fat thermogenesis[17–20].

Received 30 November 1981; accepted 31 March 1982.

1. Bray, G. A. & York, D. A. *Physiol. Rev.* **51**, 598–646 (1971).
2. Bray, G. A. & York, D. A. *Physiol. Rev.* **59**, 719–809 (1979).
3. Dubuc, P. U. *Am. J. Physiol.* **230**, 1474–1479 (1976).
4. Thurlby, P. L. & Trayhurn, P. *Br. J. Nutr.* **42**, 377–385 (1979).
5. Boissonneault, G. A., Hornshuh, M. J., Simons, J. W., Romsos, D. R. & Leveille, G. A. *Proc. Soc. exp. Biol. Med.* **157**, 402–406 (1978).
6. Thurlby, P. L. & Trayhurn, P. *Br. J. Nutr.* **39**, 397–402 (1978).
7. Kaplan, M. L. & Leveille, G. A. *Proc. Soc. exp. Biol. Med.* **143**, 925–928 (1973).
8. Kaplan, M. L. & Leveille, G. A. *Am. J. Physiol.* **227**, 912–915 (1974).
9. Trayhurn, P., Thurlby, P. L. & James, W. P. T. *Nature* **266**, 60–62 (1977).
10. Trayhurn, P. & James, W. P. T. *Pflügers Arch. ges. Physiol.* **373**, 189–193 (1978).
11. Clark, L. D. & Gray, P. E. *Biol. Psychiat.* **4**, 247–250 (1972).
12. Ohtake, M., Bray, G. A. & Azukizawa, M. *Am. J. Physiol.* **233**, R110–115 (1977).
13. York, D. A., Otto, W. & Taylor, T. G. *Comp. Biochem. Physiol.* **59B**, 59–65 (1978).
14. van der Kroon, P. H. W. & Speijers, G. J. A. *Metabolism* **33**, 1–3 (1979).
15. Thenen, S. W. & Carr, R. H. *J. Nutr.* **110**, 189–199 (1980).
16. Kaplan, M. L. & Leveille, G. A. *Am. J. Physiol.* **240**, E101–107 (1981).
17. Macdonald, I. A. & Stock, M. J. *Nutr. Metab.* **23**, 250–255 (1979).
18. Hogan, S. & Himms-Hagen, J. *Am. J. Physiol.* **239**, E301–309 (1980).
19. Thurlby, P. L. & Trayhurn, P. *Pflügers Arch. ges. Physiol.* **385**, 193–201 (1980).
20. Romsos, D. R. *Fedn Proc.* **40**, 2524–2529 (1981).

The rigidity of fish and patterns of lateral line stimulation

E. J. Denton & J. A. B. Gray

Marine Biological Association of the United Kingdom, The Laboratory, Citadel Hill, Plymouth PL1 2PB, UK

Various functions have been attributed to the lateral line organs of fish (and amphibia), including those of detecting touch and sound (both near- and far-field) and flow past a swimming fish[1–4]. As the lives of fish and the structures of their lateral lines vary greatly, lateral line function almost certainly varies between species and is probably not simple even for one animal. It is, however, generally agreed that lateral line neuromasts are excited by liquid within the canal moving relative to the canal walls[4–6]. For the sprat, such movements are proportional to local differences in motion between the fish and the surrounding seawater[7] and a similar situation must exist for other fish. Here we describe the motions of fish and seawater at various positions around vibrating sources. We show that fish are rigid longitudinally, and thus local differential movements between fish and seawater occur. Predictions based on these results suggest that when a fish is close to a source of vibration, for example, to a neighbouring fish, the amplitudes, signs and patterns of stimulation along the lateral line system change in a striking way with the position of the fish relative to the source.

Freshly killed fish (*Sprattus sprattus*, *Clupea harengus*, *Ammodytes lanceolatus* and *Trachurus trachurus*) were suspended by fine threads in a seawater tank at various distances and bearings from a vibrating source and at various angles with

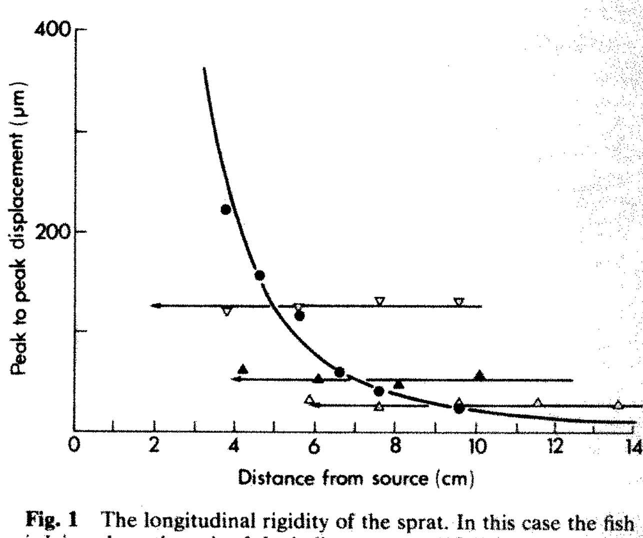


Fig. 1 The longitudinal rigidity of the sprat. In this case the fish is lying along the axis of the bellows source (12 Hz), with its head towards the source. Ordinate, peak-to-peak displacement; abscissa, distance from source. Three positions of the fish are shown; arrows indicate the snout. Triangles indicate observed fish displacement. ●, Observed water displacement; the curve shown is the best fit of a power function.

respect to a radius to the source. *S. sprattus* and *C. harengus* have lateral lines on the head only, whereas *A. lanceolatus* and *T. trachurus* have them on the head and body. The sound sources, a pulsating bellows and a vibrating sphere, were driven sinusoidally between 12 and 40 Hz. The resulting displacements of fish and water were observed with a dissecting microscope by measuring the lengths of bright lines made by the movements of small light-scattering particles in the skin of the fish and in the water. The corresponding pressures were measured with a ceramic transducer.

In any given situation, the longitudinal components of displacement of the fish (for all the species) were constant at all points, which demonstrates that these fish behave as rigid bodies longitudinally. For the small displacements used, the fish could bend freely in directions perpendicular to the vertebral column so that the amplitude of motion across the fish could differ greatly at different points along its length.

Figure 1 shows, for example, water and fish movements for a sprat with its long axis in the axis of vibration of the bellows source, at various distances from the source. The movement of the fish could generally be predicted reasonably well by assuming it to be a rigid cylinder driven by the pressure difference between its ends. While always behaving as a rigid body longitudinally, this simple method of calculation proved inadequate when the distance from the vibrator was small compared with the length of the fish.

Further studies were done using a vibrating sphere, a source more representative of a beating tail. Measurements of the amplitudes and directions of displacements and of pressures around the vibrating sphere in a seawater tank agreed well with the predictions of the near field terms of the equations given by Harris[8]. The pressures associated with such vibratory sources were much larger than we had wrongly thought[6]. We also measured movements of fish suspended at various positions with respect to the source. The pressures at the head and base of the tail were calculated and the expected motion of the fish derived from the pressure difference and the fish's length and density. Table 1 compares calculated and experimental values for the longitudinal displacements of a fish for several positions with respect to the source. As the agreement was good, it seemed reasonable to calculate displacements of the fish for other positions and also the local differences in movement between the fish and the seawater (the effective stimuli to the lateral line). Figure 2 shows such differences and how they change along the length of the fish for a source whose frequency and amplitude of motion and size are similar to those of the tail of a swimming sprat. Typically, there were one or two positions along the length of the fish at which there was no such difference. For these positions there would be no local stimulation of the lateral line. The polarity of the lateral line stimuli differ for different parts of a fish. This is probably of importance because some sensory cells in each neuromast are excited by displacements of the cupula towards the fish's head and others by displacements towards the tail. Figure 2 also gives values of pressure, longitudinal acceleration and transverse acceleration calculated for the position of the ear; in clupeids the ear contains the pressure[7] and acceleration detectors.

Table 1 Movement of fish in the field of a vibrating sphere (16 Hz)

Fish no.	r (m)	θ (deg)	ψ (deg)	ξ_{exp} (μm)	ξ_{calc} (μm)
1	0.058	1	181	90	110
2	0.087	68	132	45	31
3	0.085	66	0	52	52
4	0.104	26	85	7	7
			Mean	48.5	50

r, θ, Polar coordinates of the snout from the centre of the source (zero angle in the axis of motion). ψ, Angle of the fish to the axis of motion of the vibrating sphere. ξ, Peak-to-peak displacement of the fish along its longitudinal axis: exp, experimental; calc, calculated.

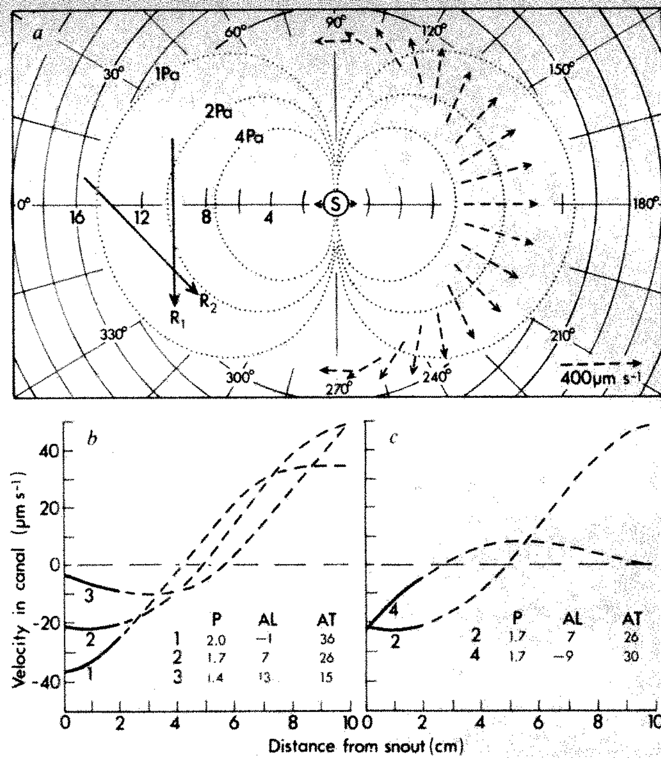


Fig. 2 *a*, Sound field in relation to the fish positions considered at the bottom of the figure. S, source, a sphere of radius 0.75 cm vibrating at 16 Hz with peak-to-peak displacements of 1.0 cm. R, 'receiving' fish, a 10 cm-long sprat, about one fish length from S. The scale in cm is shown along the 0°–180° axis. The fish are indicated by solid black lines (R_1 and R_2) with arrowheads at the snout. Dotted lines are isobars; broken lines with arrowheads show the directions and amplitudes of particle velocities in the medium. *b*, *c*, The abscissae show the distance from the snout along the length of the fish. Ordinates, velocities of the liquid in the lateral line canal (relative to the walls of the canal). The tabulated values are all peak-to-peak; P, pressure (Pa); AL, longitudinal acceleration of fish (mm s^{-2}); AT, transverse acceleration. The signs indicate the polarities during one half-cycle. The solid line shows the effects for the full extent of the sprat lateral line. Solid plus broken line indicates the relative effects for a lateral line extending to the tail. *b*, Effect of R passing S along the line of R_1 in *a* (mimicking a fish changing station in a shoal): (1) fish 2 cm behind position labelled R_1; (2) fish in position R_1; (3) fish 2 cm ahead of position R_1. *c*, Effect of R changing its angle by 45°; (2) fish in position R_1; (4) fish in position R_2.

For the sprat we know the relationships between direction, frequency and amplitude of seawater movements outside the fish and movements within the lateral line canals[7]. Making an informed guess at the sensory thresholds from the results of various experiments (refs 7, 9 and some unpublished results), we estimate that the sprat lateral line system excited in this way could detect a neighbouring fish in a shoal at distances of up to a few fish lengths.

These experiments relate to situations that are simpler than the natural ones. They do show, however, for both a vibrating and a pulsating source, that the stimulation of lateral line sense organs (by a neighbouring source) falls more rapidly with distance and changes more dramatically with position than might be expected. The stimulations of the lateral line, otolith[10] and pressure-sensitive sense organs all change in different ways on moving around a vibrating (Fig. 2) or pulsating source. In this array of sense organs, those of the lateral line, in which the pattern of stimulus changes greatly with position, are well suited to having a major sensory role in such activities as the schooling of sprat and saithe[11] or, in the case of rays, searching for prey.

When the distance of a source is great relative to the length of a fish (this can happen well within the near field), the motion of the water at the head of the fish will be very similar to that at the tail. The differences in motion between fish and water

will thus be very small, therefore the stimulation of the lateral line by the mechanism discussed here will be of litle consequence. This means that lateral line systems should be disturbed little by most of the background noises in the sea[4,6,12].

We thank Mr C. R. Griffiths for generous help. J.A.B.G. is a member of the external scientific staff of the MRC. The Association is grant-aided by the NERC.

Received 17 March; accepted 26 April 1982.

1. Harris, G. G. & Van Bergeijk, W. A. *J. acoust. Soc. Am.* **34**, 1831–1841 (1962).
2. Dijkgraaf, S. *Biol. Rev.* **38**, 51–105 (1963).
3. Kuiper, J. W. *Lateral Line Detectors* (ed. Cahn, P.) 105–121 (Indiana University Press, 1967).
4. Sand, O. *Hearing and Sound Communication in Fishes* (eds Tavolga, W. N., Popper, A. N. & Fay, R. R.) 459–480 (Springer, New York, 1981).
5. Sand, A. *Proc. R. Soc.* **B123**, 472–495 (1937).
6. Denton, E. J., Gray, J. A. B. & Blaxter, J. H. S. *J. mar. biol. Ass. U.K.* **59**, 27–47 (1979).
7. Blaxter, J. H. S., Denton, E. J. & Gray, J. A. B. *Hearing and Sound Communication in Fishes* (eds Tavolga, W. N., Popper, A. N. & Fay, R. R.) 39–59 (Springer, New York, 1981).
8. Harris, G. G. *Marine Bioacoustics* (ed. Tavolga, W. N.) 233–247 (Pergamon, Oxford, 1964).
9. Enger, D. S. *Comp. Biochem. Physiol.* **22**, 527–538 (1967).
10. Hawkins, A. D. & Horner, K. *Sound Communication in Fishes* (eds Tavolga, W. N., Popper, A. N. & Fay, R. R.) 311–328 (Springer, New York, 1981).
11. Partridge, B. L. & Pitcher, T. J. *J. comp. Physiol.* **135**, 315–325 (1980).
12. Urick, R. J. *Principles of Underwater Sound* (McGraw-Hill, New York, 1967).

Place navigation impaired in rats with hippocampal lesions

R. G. M. Morris*, P. Garrud*, J. N. P. Rawlins† & J. O'Keefe‡

* Psychological Laboratory, University of St Andrews, St Andrews, Fife KY16 9JU, UK
† Department of Experimental Psychology, University of Oxford, Oxford OX1 2JD, UK
‡ MRC Cerebral Mechanisms Unit, University College, University of London, London WC1E 6BT, UK

Electrophysiological studies have shown that single cells in the hippocampus respond during spatial learning and exploration[1–4], some firing only when animals enter specific and restricted areas of a familiar environment. Deficits in spatial learning and memory are found after lesions of the hippocampus and its extrinsic fibre connections[5,6] following damage to the medial septal nucleus which successfully disrupts the hippocampal theta rhythm[7], and in senescent rats which also show a correlated reduction in synaptic enhancement on the perforant path input to the hippocampus[8]. We now report, using a novel behavioural procedure requiring search for a hidden goal, that, in addition to a spatial discrimination impairment, total hippocampal lesions also cause a profound and lasting place-navigational impairment that can be dissociated from correlated motor, motivational and reinforcement aspects of the procedure.

If rats are placed in a large circular pool of opaque water, they will quickly learn to escape by finding and climbing on to a small platform hidden beneath the water surface, provided it remains in a fixed location over a series of trials[9]. They cannot learn to find it when its position varies randomly from trial to trial. Although they can never see, hear or smell the platform, rats require only a few trials in order to learn to swim directly towards it, using the shortest route, even from a novel starting place. That is, the rats learn not only to recognize the vicinity of the safe place when they reach it, but also to swim towards it from a distance despite the absence of cues from the platform itself. The deleterious effects of cue–response separation apparent in visual discrimination[10,11] do not, in this case, prevent extremely rapid learning. By comparing the performance of normal and brain-lesioned animals in these conditions with that shown when a fixed but visible platform was used, we have examined the role of the hippocampus in simple navigation.

Female Lister rats ($n = 31$) were subjected to the following procedures: total hippocampal lesions ($n = 10$), superficial cortical lesions ($n = 13$), sham surgery ($n = 4$) or no surgery ($n = 4$). The rats were placed, under pentobarbitone anaesthesia, into a special adjustable head-holder[12]. Animals in the hippocampal lesion group had holes drilled in their skulls, and a small amount of neocortex overlying the hippocampus, and the entire dorsal and ventral hippocampus were removed by aspiration. Operated control animals had comparable lesions in the neocortex but showed no hippocampal damage. Sham-operated control animals had burr holes drilled in their skulls but suffered no brain damage. On completion of the behavioural procedures, conventional histological techniques (40 μm, gelatine-embedded sections stained with cresyl violet and solochrome cyanide) were used to verify the lesions. All rats of the hippocampal-lesion group were found to have total or near total destruction of the dorsal and ventral hippocampus, with minimal damage to adjacent structures (comparable lesions are reported in ref. 13). Analysis of the behavioural data showed no differences between the sham-operated and unoperated control groups, which were therefore combined, giving final group sizes of 10 (hippocampal), 13 (cortical) and 8 (control).

On day 1, the rats were placed in a pool of water (1.32 m diameter, 53 l at 26 ± 1 °C) and allowed to swim freely for 1 min with no opportunity for escape. On day 2, a platform was hidden in one of four locations in the middle of each cardinal quadrant (SW, NW, NE and SE), 0.33 m from the side walls. Different locations were used for different rats. The platform, made of clear perspex, was hidden by adding 2.3 l of milk to the water and arranging for its top surface, 8 cm in diameter, to be 1 cm below the water level. A second platform, 2 cm taller and protruding visibly out of the water, was used at a later stage of training; its top surface was indented such that it contained within its circumference a 1 cm layer of water. Thus, the reinforcement afforded by escape on to the two platforms was equated. The rat's task throughout the training procedures, which continued for 8 days, was to find and escape on to the platforms. Only one platform was used at a given stage of training, and it was always in a fixed position in the pool on a given day. Thus the two tasks, which we shall call place-navigation and cue-navigation, respectively, involved the same motor movements (swimming), motivation and reinforcement (escape from water), but differed specifically and uniquely with respect to whether or not the rat was required to learn the platform's position in relation to the varied distal room cues.

All rats swam effectively using the characteristic adult swimming posture[14]. The times taken to escape from the water during the three successive phases of the experiment are shown in Fig. 1. The normal and cortical-lesion groups learned to escape rapidly from the water with stable terminal acquisition latencies of < 8 s. The hippocampal-lesion group showed a highly significant impairment in the place-navigation task (trials 1–28) when the hidden platform was used. However, this impairment declined dramatically and disappeared when the visible platform was used (trials 30–41), this platform having been placed diagonally opposite to the earlier training location (that is, NW for a rat trained previously to find the hidden platform at SE). The place-navigation impairment reappeared when training was continued with the hidden platform (trials 43–50) even though it remained in the same position that the visible platform had occupied in the preceding phase of training.

Detailed analysis of the behavioural performance of each group and the results of two transfer tests provide new insights into the nature and magnitude of the deficit after hippocampal lesions. First, the hippocampal-lesion animals did improve during training but never escaped faster than normal animals searching for a hidden platform that was moved around randomly from place to place on successive trials (see Fig. 1, horizontal broken line, taken from ref. 9). Second, analysis of the paths taken by all rats on trial 28, transcribed from videotape recordings, showed that the hippocampal-lesion animals took longer and more circuitous routes to find the hidden platform

0028-0836/82/250681—03$01.00

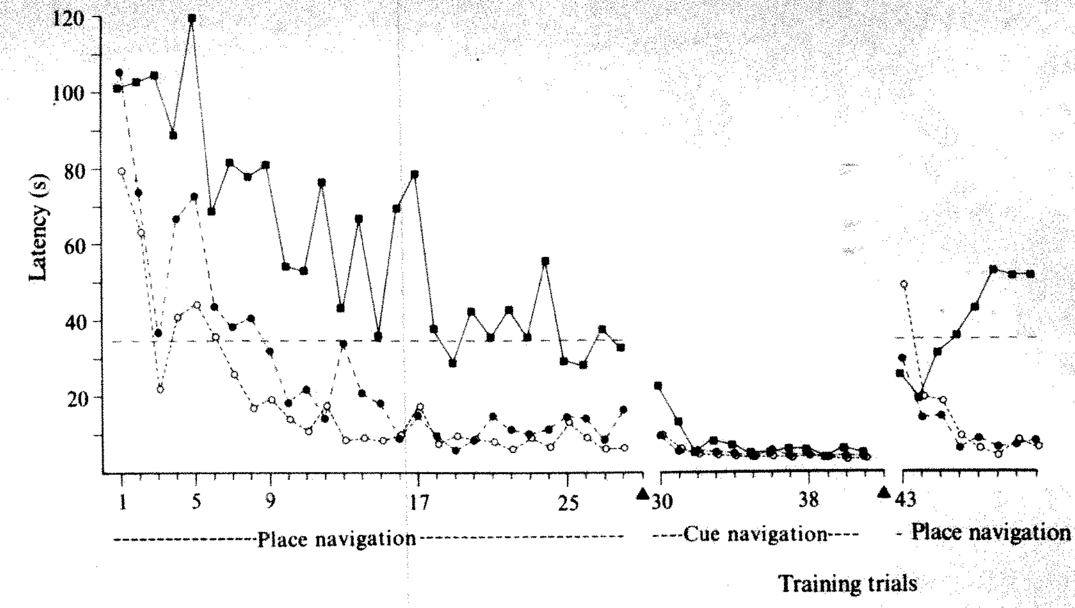


Fig. 1 Mean latency of escape (s) for the 50 trials of the experiment. ■, Hippocampal lesion; ●, cortical lesion; ○, control;. The trial number of the first of each daily set of trials is shown on the abscissa. The two transfer tests are indicated by solid triangles. To avoid problems of heterogeneity of variance, the successive phases of the experiment were analysed separately. The horizontal broken line (trials 1–28 and 43–50) at 34.5 s corresponds to the best performance shown by a group of normal rats trained to search in 20 trials for a hidden platform that was moved randomly from one place to another over successive trials (data taken from ref. 9). Place navigation (trials 1–28): unweighted means (unequal n) analysis of variance revealed significant effects of group ($F=23.7$, d.f. $=2/28$; $P<0.0001$), trials ($F=17.8$, d.f. $=23/667$; $P<0.0001$) and lesion × trials ($F=1.5$, d.f. $=54/756$; $P<0.02$). Subsequent orthogonal comparisons showed that the deficit was restricted to the hippocampal-lesion rats (hippocampal versus cortical + control, $P<0.0001$; cortical versus control, $P>0.10$). Cue navigation (trials 30–41): terminal escape latencies (trial 41) were 5.0, 3.3 and 2.8 s for the hippocampal-lesion, cortical-lesion and control groups respectively, corresponding to declines relative to trial 28 of 45.9, 7.9 and 3.7 s). Analysis of variance of all 12 trials revealed a small residual impairment in the hippocampal-lesion groups ($F=5.4$, d.f. $=2/28$; $P<0.02$). Return to place navigation (trials 43–50): analysis of variance showed a highly significant effect of groups ($F=12.2$, d.f. $=2/28$; $P<0.0002$) and a lesion × trials interaction ($F=3.6$, d.f. $=14/196$; $P<0.0001$). The apparent gradual impairment of performance in the hippocampal-lesion group was caused by a slowing of swimming speed over trials as the core temperature of the rats fell slightly (from 37 °C to ~35 °C).

(Fig. 2). The directional heading of the hippocampal-lesion rats when they set off from their starting position on trial 28 was no more likely to be towards the platform than in any other possible direction. These results imply that hippocampal-lesion rats can learn some sort of escape strategy (for example, that escape is possible) but are substantially poorer at learning where the hidden platform is located and, unlike normal and cortical-lesion animals, will never learn to swim towards it from a distance.

The magnitude of this place-navigational deficit was assessed in two separate transfer tests conducted on trials 29 and 42, immediately after the four daily trials of days 6 and 8. For transfer test A, the hidden platform was first removed from the apparatus, then the rats were placed in the pool for 60 s with no opportunity for escape, and their movements observed. The results were striking. Control and cortical-lesion rats swam to and persistently across the former platform location whereas the hippocampal-lesion rats did not. The hippocampal-lesion rats did not merely swim around the side walls. To demonstrate this, annuli were marked on the video screen indicating the exact surface area and former positions of the platforms in each of the four cardinal quadrants. The total number of annuli which an individual rat passed through during the 60-s test was 7.6, 6.8 and 8.6 for the hippocampal-lesion, cortical-lesion and control groups, respectively ($F<1$). The groups were distinguished by which annuli they passed through: an individual hippocampal-lesion rat was no more likely to pass through the annulus marking the platform position used during training than one in any other quadrant (Fig. 3a). We observed no tendency on the part of the hippocampal-lesion rats to remain in the vicinity of the training annulus once they had eventually reached it (compare with ref. 7). Thus the deficit produced by hippocampal lesions was total. Furthermore, with respect to the lack of spatial bias revealed in the annulus measure, the deficit was apparent in all 10 rats of the experimental group.

Our interpretation of these findings is that, whatever their other effects[15–19], hippocampal lesions do cause a profound and lasting place-navigational impairment. It could be argued, however, that while matched for motor requirements, motivation and reinforcement, the place- and cue-navigational tasks are not matched for task complexity. Perhaps hippocampal-lesion animals perform poorly on the spatial task because it is complex (albeit a task learned by normal animals in less than 10 trials), and perform better on the visible platform task because it is easier, rather than because the spatial component is then redundant. If this is the basis of the dissociation of effects in the two tasks, then at least some spatial bias should be shown by some of the hippocampal-lesion animals in a transfer test conducted after training on the ostensibly easier visible platform task. Transfer test B, conducted immediately after trial 41 on day 8, examined this possibility. In trial 41 itself, there was no significant difference in the latency, path-length or directionality of escape behaviour across groups ($F_s<1$), all animals escaping rapidly by means of short, direct paths to

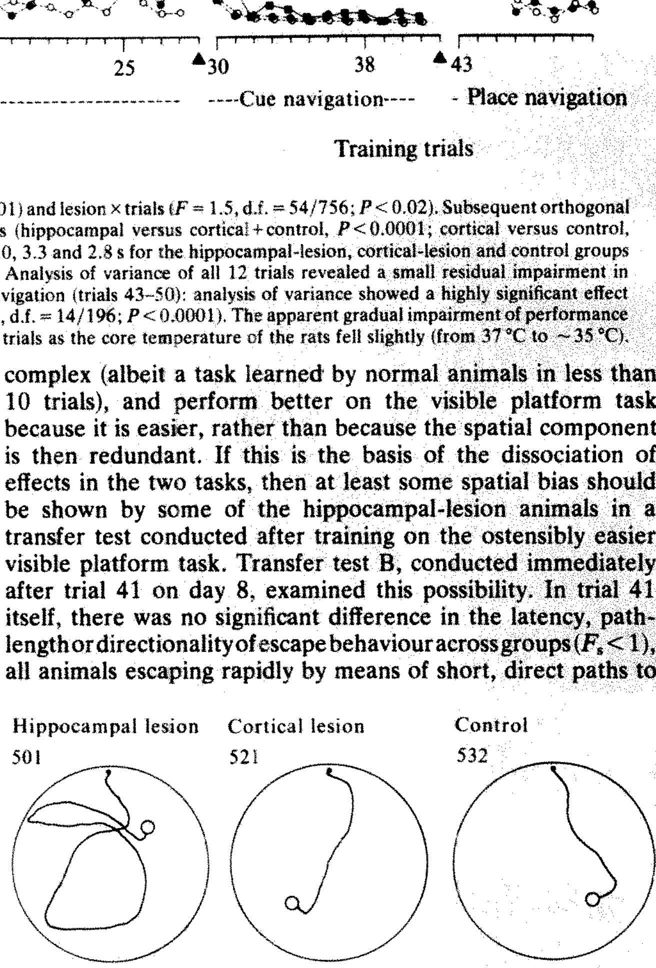


Fig. 2 The actual path of the median rat (defined in terms of path length) in each group on trial 28 just before the first transfer test. The rats were observed using a video camera placed above the pool. One experimenter (P.G.) sat concealed in one corner of the room and monitored the rat's movement on a VTR recorder. The second experimenter (R.M.) removed each rat from its home cage in an adjacent room and placed it in the pool. The pool was open to the room which included a door, a window, and brightly and darkly lit walls. The paths taken by the rats in escaping were transcribed from the videotape and measured. Path lengths: the hippocampal-lesion rats took 4.66 ± 0.86 m to reach the platform, whereas the cortical-lesion and control rats took 2.35 ± 0.98 and 1.20 ± 0.34 m, respectively. Analysis of variance showed that these path lengths differed significantly ($F=4.23$, d.f. $=2/28$; $P<0.025$). Subsequent orthogonal comparisons showed that the hippocampal-lesion group took significantly longer paths than both the cortical-lesion and control groups ($P<0.001$), which in turn did not differ significantly from each other ($P>0.10$). Directionality: the accuracy of the approach to the platform was analysed as follows. We measured the angle subtending a tangent to the rat's path at a point 0.5 m from its starting position, and a line intersecting this point and the centre of the platform. This angle was $83\pm18°$ from the correct direction for the hippocampal-lesion group, whereas for the cortical-lesion and control groups, the angle was $34\pm13°$ and $38\pm17°$, respectively (Kruskal–Wallis, $H=5.87$, d.f. $=2$; $P<0.025$, one-tailed).

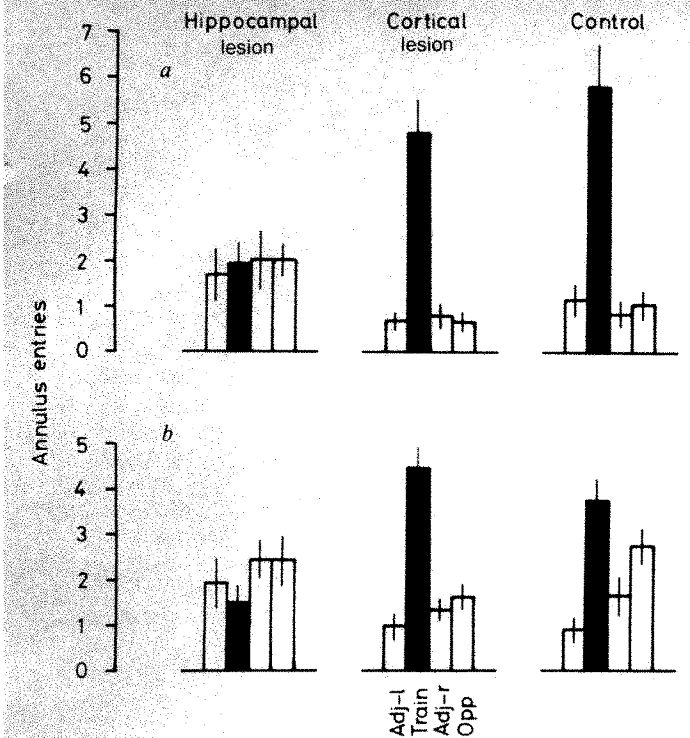


Fig. 3 Mean crossings of each of the annuli (±1 s.e.) marking the former platform positions during *a*, transfer test A (after place-navigation training) and *b* transfer test B (after cue-navigation training). The data have been categorized for each animal into crossings of the training location (Train), and of the annulus in the adjacent quadrant to the left (Adj-l; viewed from above), the adjacent right (Adj-r), and opposite quadrant (Opp). Note that the hippocampal-lesion rats were no more likely to pass through the annulus marking the training position than any other, in both transfer tests. Analyses of variance showed a highly significant groups x position effect in both transfer test A ($F = 10.1$, d.f. $= 6/84$; $P < 0.0001$) and transfer test B ($F = 9.5$, d.f. $= 6/84$; $P < 0.0001$).

the visible platform. However, in the transfer test conducted 30 s later, only the control and cortical lesion groups searched in the vicinity of the now absent but previously visible platform (Fig. 3*b*). All 10 animals in the hippocampal-lesion group showed no spatial bias. Thus even if the improvement by the hippocampal-lesion group in the visible platform phase of training was due to the simplicity of the task, this improvement was not accompanied by any spatial learning.

The procedures used here provide a new approach to analysing the brain mechanisms of spatial localization. The results show that hippocampal lesions cause a profound and lasting impairment in place-navigation and question that aspect of the working-memory hypothesis[17] which asserts that spatial reference memory is unaffected by septo-hippocampal damage. Reference memory has been defined as those aspects of a learning procedure in which learned information may be used in every trial of training rather than for just a single trial. The present procedure using a fixed platform position for 28 trials, followed by different fixed position for 20 further trials, is certainly a reference-memory procedure. In the absence of separate measures of working memory in this experiment, we cannot comment further on the adequacy of that hypothesis. However, we suspect that claims about the integrity of spatial reference memory after more restricted fimbria-fornix lesions and extensive preoperative training[20] may provide a misleading picture of normal hippocampal function.

The present results show that normal rats can navigate in an appropriate direction towards a hidden object; they do not merely recognize a place when they reach it. Whether this type of learning involves or is different from conventional associative learning deserves further scrutiny. But given that place units detected so far in the hippocampus[1–4] respond only with respect to places in which the rat is presently situated as opposed to places to which it intends to go, these results pose a challenge for electrophysiologists attempting to explain the neural mechanisms by which the hippocampus processes spatial information.

We thank A. Ainsworth for advice about histology, C. Barman for his assistance in constructing the apparatus, and numerous colleagues for their comments on earlier drafts of the paper. This research was supported in part by MRC research grant G979/265/N.

Received 18 January; accepted 14 April 1982.

1. O'Keefe, J. *Brain Res.* **34**, 171–175 (1971).
2. Best, P. J. & Ranck, J. B. Jr *Soc. Neurosci. Abstr.* **1**, 538 (1975).
3. O'Keefe, J. *Expl Neurol.* **51**, 78–109 (1975).
4. Olton, D. S., Branch, M. & Best, P. J. *Expl Neurol.* **58**, 387–409 (1978).
5. O'Keefe, J., Nadel, L., Keigthly, S & Kill, D. *Expl Neurol.* **58**, 387–409 (1978).
6. Olton, D. S., Walker, J. A. & Gage, F. A. *Brain Res.* **139**, 295–308 (1978).
7. Winson, J. *Science* **201**, 160–163 (1978).
8. Barnes, C. A. *J. comp. physiol. Psychol.* **93**, 74–104 (1979).
9. Morris, R. G. M. *Learn Motiv.* **12**, 239–260 (1981).
10. Cowey, A. in *Discrimination* (ed. Weiskrantz, L.) 189–238 (Cambridge University Press, 1968).
11. Milner, A. D., Goodale, M. A. & Morton, M. *J. comp. physiol. Psychol.* **93**, 1015–1023 (1979).
12. Rawlins, J. N. P. & Bennett, J. *Physiol. Behav.* **24**, 415–416 (1980).
13. Rawlins, J. N. P., Feldon, J. & Gray, J. A. *Expl Brain. Res.* **38**, 49–63 (1980).
14. Schapiro, S., Salas, M. & Vukovich, K. *Science*, **168**, 147–149 (1970).
15. Gaffan, D. *J. comp. physiol. Psychol.* **86**, 1100–1109 (1974).
16. Mishkin, M. *Nature* **273**, 297–298 (1978).
17. Olton, D. S., Becker, J. T. & Handelmann, G. E. *Behav. Brain Sci.* **2**, 313–365 (1979).
18. Gray, J. A. *The Neuropsychology of Anxiety* (Oxford University Press, in the press).
19. Weiskrantz, L. *Proc. R. Soc.* B (in the press).
20. Olton, D. S. & Papas, B. *Neuropsychologia* **17**, 669–682 (1979).

Evidence for dendritic competition in the developing retina

V. H. Perry & R. Linden

Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK

At present little is known of the rules regulating dendritic morphology. Several studies have demonstrated that the shape of the dendritic tree depends on its afferent supply[1,2]. The ganglion cells of the retina provide a particularly useful cell type for the study of neurone development as they develop independently of afferents from other brain regions. If the ganglion cells alone are destroyed in a small patch of the developing retina, it is possible to examine how the absence of neighbouring neurones of the same type influences the development of the ganglion cells around the depleted area. The development of the normal laminar pattern of the retina is not disturbed by the loss of these cells[3]. We show here that the dendrites of ganglion cells around the depleted area are preferentially directed towards this region. The orientation of ganglion cell dendrites is strongly influenced by neighbouring cells and we suggest that during normal development, dendrites compete for their afferents.

Experiments were performed on 11 hooded Lister rats. On the day of birth, the rats were anaesthetized by hypothermia and a small lesion was made in the temporal retina of one eye using a fine 28-gauge needle passed through the sclera approximately half-way between the optic disk and the limbus. After 2–3 months, the animal were anaesthetized with an intraperitoneal injection of 3.0 ml per kg of chlor-nembutal (2.1 g of chloral hydrate + 0.5 g of sodium pentobarbital in 50 ml of 0.9% saline). A series of six injections of 0.15–0.25 µl of horseradish peroxidase (HRP; Boehringer) (50% w/v in 2% dimethyl sulphoxide) were made stereotaxically into the optic tract, using a 1-µl Hamilton syringe. The animals were killed painlessly after 24 h, perfused with 0.9% saline and the eyes removed. The retinae were prepared as whole mounts[4], and reacted in a modified Hanker–Yates solution[5]. After washing in 0.1 M phosphate buffer (pH 7.2) for several hours, each retina was transferred to 50 ml solution of 0.1 M sodium

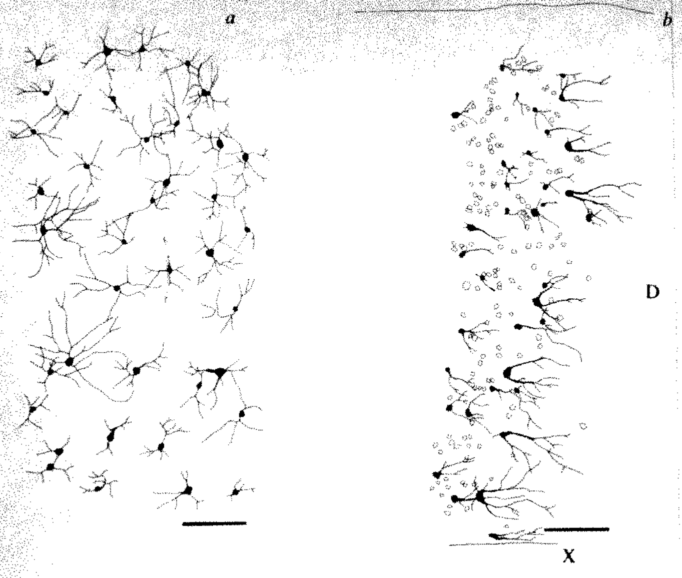


Fig. 1 *a*, Drawing of the cell bodies and proximal dendrites of ganglion cells filled with HRP. The HRP reaction product did not fill the entire dendritic tree. The cells are located ~3 mm nasal to the optic disk in a retina with a lesion located temporal to the optic nerve head. They were drawn to give an approximately evenly spaced distribution; other labelled neurones are not shown. *b*, Drawing of the cells along the edge of a region devoid of ganglion cells (D). The site of entry of the needle is marked by X; the line at the top of the figure indicates the periphery of the retina. The cell bodies of some faintly filled cells are drawn in outline to indicate the border of the region devoid of ganglion cells. The dendrites and cell bodies of well-filled cells are drawn completely. Reproduced from an original drawing made at ×400. Scale bar, 250 μm.

cacodylate buffer (*p*H 5.1) containing (mg) ammonium nickel sulphate (200), cobalt chloride (300), *p*-phenylenediamine (25) and catechol (50). They were incubated in this solution for 15 min, then washed in phosphate buffer for 3–5 min and transferred to a fresh solution of 50 ml of 0.1 M sodium cacodylate buffer (*p*H 5.1), containing *p*-phenylenediamine (25 mg), catechol (50 mg) and 1 drop of 30% hydrogen peroxide. After incubation for 15 min, the retinae were again washed in phosphate buffer and mounted on a gelatinized slide. The retinae from several normal animals (that is, having no lesions) injected with HRP were treated in the same way as those from the experimental animals and served as controls. The retinae from three of the experimental animals were fixed in 10% formol-saline, prepared as whole mounts and stained with cresyl violet. In addition, retinae from animals in which one optic nerve had been sectioned at birth, were stained with a Nissl stain either as whole mounts or in vertical sections[6].

In the retinae of the normal rats, many ganglion cells were filled sufficiently with HRP to reveal their proximal dendrites. It proved difficult to follow the dendrites for a long distance before they became lost in the dense network of the inner plexiform layer, except for a few cells where a large part of the dendritic tree was intensely stained and resembled a Golgi-stained neurone. As has been previously reported for the rat and other animal species[7,8], most ganglion cells have circular or elliptical dendritic fields but some have dendrites predominantly on one side of the cell body. Within any given patch of retina, however, such cells showed no consistent asymmetry (see Fig. 1). The only exceptions are the outermost rim of the retina, along which ganglion cell dendrites tend to be oriented towards the ora serrata, and the region of the major blood vessels where the cell bodies have been displaced laterally away from their underlying afferents.

In the experimental retinae, the lesion site was readily identified because the ganglion cell layer peripheral to it was almost totally lacking in ganglion cells. The needle cut the axons of the cells peripheral to the lesion in a wedge-shaped segment of the retina. These ganglion cells had degenerated, in agreement with previous studies which showed that neonatal ganglion cells do not survive transection of their axons[6]. At the lesion site there was some disruption of the retinal layers but in the peripheral region lacking ganglion cells, the laminar organization was normal for except for some shrinkage of the inner plexiform layer. The thickness of the retinal layers was measured in animals in which one optic nerve had been sectioned at birth, and which were kept alive for up to 11 months. The inner plexiform layer had shrunk by as much as 50% whereas the other layers showed only 3% shrinkage compared with normal. The area of the retinae with the optic nerve sectioned was similar to the normal retinae and the loss of the ganglion cells did not disrupt the formation of the normal lamina pattern[3,6]. We therefore conclude that despite the loss of ganglion cells, the bipolar and amacrine cells which constitute the afferent supply of the depleted area, remained largely intact.

In the HRP-stained retinae, it was apparent that the dendrites of the ganglion cells adjoining the lesion site had their dendrites directed almost exclusively towards the region lacking ganglion cells. This was true irrespective of cell soma size, and included the occasional displaced ganglion cell in the inner nuclear layer (Figs 1*b*, 2). This effect on the orientation of the dendrites can be seen all around the depleted area, irrespective of the shape of this area (Fig. 3) and is therefore probably not a consequence of any simple mechanical distortion. Ganglion cells located at least 400 μm from the edge of the depleted area were found to have distorted dendrites.

The results show that the orientation of developing ganglion cell dendrites is strongly influenced by neighbouring cells. The dendritic fields of developing ganglion cells are predominantly radial[9], but in these experiments we observed cells having almost all their dendrites on one side of the cell body. We do not yet know the sequence of events which results in this dramatic reorganization of the dendritic geometry; the dendrites may be attracted to the depleted region and/or repelled by their neighbours on the other side. We suggest that during normal development, ganglion cell dendrites compete with each other for their afferent supply from bipolar and amacrine cells.

There is considerable evidence that the axon terminals of ganglion cells and other neurones compete for their postsynaptic territory during development[10,11]. In the rodent retina, many ganglion cells degenerate in the first few days postnatal[12,13]. The degeneration of ipsilaterally projecting cells can be partially prevented by removing one eye at birth, thus reducing the competition between the developing optic axon terminals[12,14]. We have recently shown[15] that after transection of one optic tract at birth, the number of cells which project ipsilaterally to the brain from the eye contralateral to the tract section is

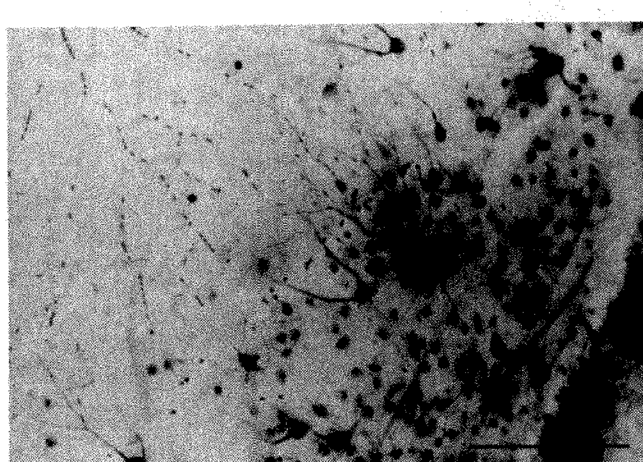


Fig. 2 Photomicrograph showing some of the ganglion cells with their dendrites directed towards a region lacking ganglion cells. Scale bar, 250 μm.

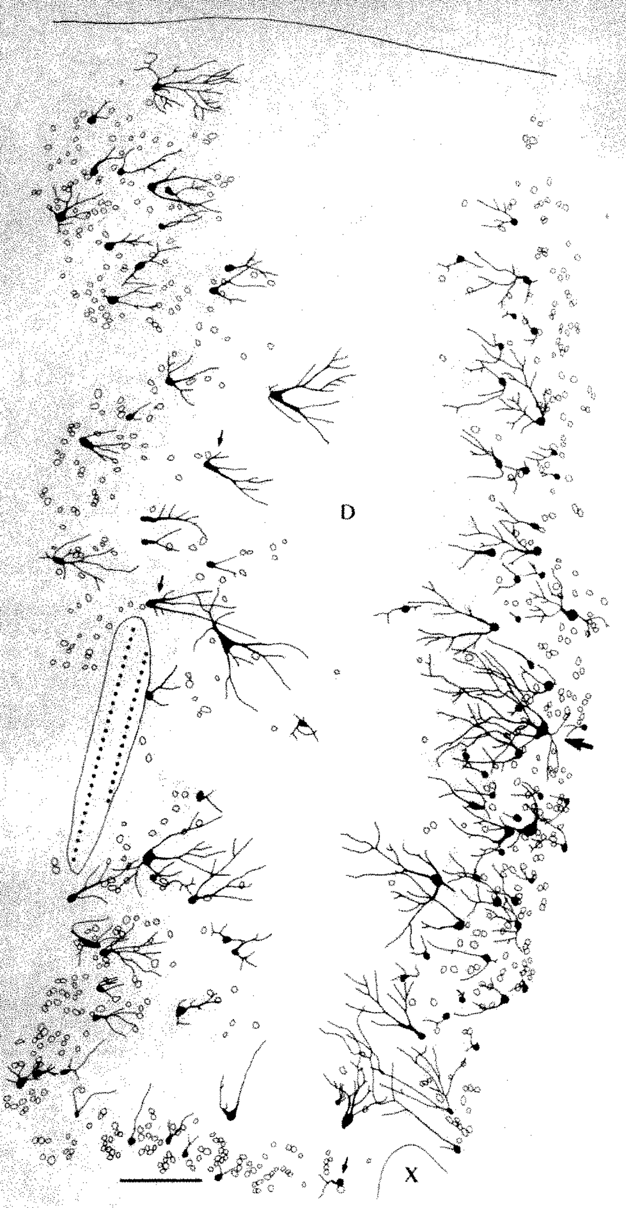


Fig. 3 Drawing of a small rectangular region of retina lacking ganglion cells, showing surrounding HRP-filled neurones. Conventions as for Fig. 1*b*. Stippling represents an area of post-mortem damage. The small arrows indicate displaced ganglion cells. The large arrow indicates the large cell illustrated in the centre of Fig. 2. Scale bar, 250 μm.

increased. This is accompanied by an increase in the area occupied by the ipsilateral terminal field. This increase in the ipsilateral projection occurs despite the fact that the crossed projection from the other eye remains intact and that the expanded projection is found away from regions receiving commisural fibres. Thus the effect is not simply due to a change in competition at the axon terminals. We suggested that this abnormal population of ganglion cells survived in the retina as a result of reduced dendritic competition since the optic tract damage induced retrograde degeneration of almost all the contralaterally projecting ganglion cells. The results presented here provide evidence for such an interaction between adjacent developing ganglion cell dendrites.

It has been shown that different retinal cell types are distributed in regular mosaics across the retina[16] and that the dendritic trees of a particular cell class are arranged so that every part of the retina is covered by the dendrites of that cell type[17], presumably as a result of local interactions[16,18]. If the population of ganglion cells in the developing retina is largely in excess of that found in the adult[12,13], a regular mosaic can be generated by dendritic competition regulating cell death. Ganglion cells which fail to make a sufficient number of contacts with their afferents would die in much the same way as neurones which fail to make the appropriate postsynaptic contacts also die[12,14]. The shape of the dendritic field of a given cell will result from the orientation of the dendrites changing in accord with available afferent supply. It remains to be seen whether other aspects of dendritic geometry, such as the length of the dendrites and frequency of branching, can also be influenced by the density of neighbouring cells and whether these interactions are limited within each given class of ganglion cell or level of branching in the inner plexiform layer.

We thank Professor A. Cowey and Dr K. A. C. Martin for their comments on the manuscript. This work was supported by the MRC. R.L. is supported by a fellowship from Conselho Nacional de Desenvolvimento Cientifico e Tecnologico (CNPq-Brazil Proc. 200.512-79 BQ).

Received 22 February; accepted 28 April 1982.

1. Bradley, P. & Berry, M. *Brain. Res.* **109**, 133–151 (1976).
2. Borges, S. & Berry, M. *Brain Res.* **112**, 141–147 (1976).
3. Eayrs, J. T. *Br. J. Ophthal.* **36**, 453–459 (1952).
4. Cowey, A. & Perry, V. H. *Expl Brain Res.* **35**, 457–464 (1979).
5. Hanker, J. S., Yates, P. E., Meltz, C. B. & Rustioni, A. J. *Histochem. J.* **9**, 789–792 (1977).
6. Perry, V. H. *Neuroscience* **6**, 931–944 (1981).
7. Perry, V. H. *Proc. R. Soc.* B **204**, 363–375 (1979).
8. Boycott, B. B. & Wässle, H. J. *J. Physiol., Lond.* **204**, 397–419 (1974).
9. Perry, V. H. & Walker, M. *Proc. R. Soc.* **B208**, 415–431 (1980).
10. Land, P. W. & Lund, R. D. *Science* **205**, 698–700 (1979).
11. LeVay, S., Wiesel, T. N. & Hubel, D. H. *J. comp. Neurol.* **191**, 1–51 (1980).
12. Sengelaub, D. R. & Finlay, B. L. *Science* **213**, 573–574 (1981).
13. Cunningham, T. J., Mohler, I. M. & Giordano, D. L. *Devl Brain. Res.* **2**, 203–215 (1981).
14. Jeffery, G. & Perry, V. H. *Dev. Brain Res.* **2**, 176–180 (1981).
15. Linden, R. & Perry, V. H. *Neuroscience* (in the press).
16. Wässle, H. & Reimann, H. J. *Proc. R. Soc.* **B200**, 441–461 (1978).
17. Wässle, H., Peichl, L. & Boycott, B. B. *Proc. R. Soc.* **B212**, 157–175 (1981).
18. Wässle, H., Peichl, L. & Boycott, B. B. *Nature* **292**, 344–345 (1981).

Incorporation of cytoplasmic vesicles into apical membrane of mammalian urinary bladder epithelium

Simon A. Lewis & José L. C. de Moura*

Department of Physiology, Yale University School of Medicine, New Haven, Connecticut 06510, USA

The mammalian urinary bladder is a distensible organ which undergoes a series of slow fillings and rapid emptyings (micturition), during a 24 h period. Recent evidence suggests that the bladder epithelium undergoes a three-phase, structural change during the expansion stage of this cycle[1,2]. First, macroscopic multicellular folds are smoothed then microscopic folds are stretched out, resulting in a flattening of the urine-facing (apical) membrane. The last and most speculative phase is an incorporation or fusion of cytoplasmic vesicles with the apical membrane, resulting in an alteration of the cell shape from cuboidal to squamous. Such fusion of cytoplasmic vesicles is interesting not only from the point of view of the bladder as a storage reservoir for the discharge of the kidneys, but also from the perspective of hormonal regulation of the ion reabsorptive processes in this epithelium[1,3]. Recently, the determination of the electrical properties of the apical membrane of mammalian urinary bladder has revealed that aldosterone (a Na^+-conserving hormone) increases both Na^+ conductance and selectivity[4]. The origin of such highly Na^+ selective channels is unknown. Here we report that mammalian urinary bladder epithelium accommodates stretching by incorporating into the apical membrane vesicles from a cytoplasmic pool, which are withdrawn to effect recovery from stretch. The vesicle membrane contains Na^+-selective channels—this has implications for the regulation of membrane transport.

* Present address: Department of Urology, Universidad Lisboa, 1699 Lisboa, Portugal.

Rabbit urinary bladder epithelium was dissected, stripped of underlying muscular coats and mounted in modified Ussing chambers ($2\ cm^2$ nominal area) designed to eliminate edge damage as previously described[1]. Both sides of the epithelium were bathed at all times with identical solutions. Open circuit potential (T_T), trans-epithelial resistance (R_T) and short-circuit current (I_{sc}, a direct measure of Na^+ transport) were continuously monitored and stored on a laboratory computer.

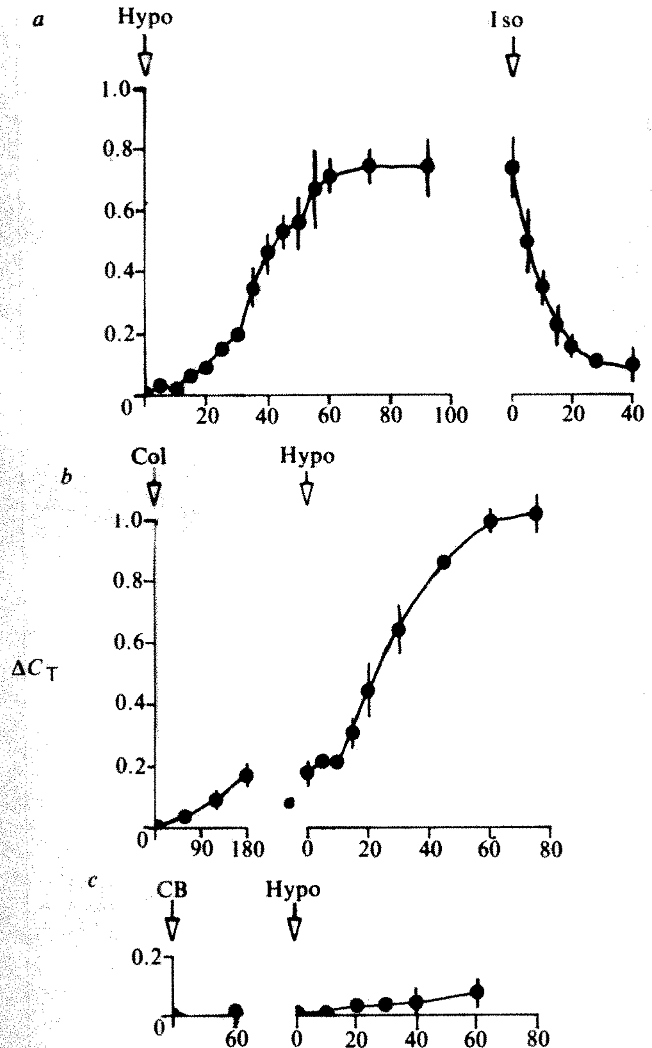


Fig. 1 Increase in normalized capacitance (ΔC_T) as a function of time after both mucosal and serosal solutions were replaced with a solution of half osmolarity. Capacitance was determined by passing a square current pulse across the epithelium, digitizing the voltage response and then fitting the voltage response by the sum of two exponentials. The best fit capacitance values (C_1 and C_2) are related to the actual apical (C_a) and basolateral (C_{bl}) capacitances by

$$C_T = \frac{C_1 C_2}{C_1 + C_2} = \frac{C_a C_{bl}}{C_a + C_{bl}}.$$

As $C_a : C_{bl}$ is 1:5 (ref. 5), then C_T is an underestimate of C_a. If a change in C_T is produced by an increase in only C_{bl} then C_T can change by a maximum of only 20% (that is, C_{bl} becomes infinite). The values shown in the figure indicate that the increase in C_T was almost completely accounted for by an increase in C_a, not C_{bl}. Capacitance was normalized by subtracting the measured capacitance from the zero time capacitance and dividing this difference by the zero time capacitance. Vertical bars represent the s.e.m. *a*, Time course of change in C_T after changing the bathing solutions (at $t = 0$) to half osmolarity (Hypo) and back to isosmotic solutions (Iso). Actual capacitance was $1.14 \pm 0.07\ \mu F\ cm^{-2}$ ($n = 8$) at $t = 0$ and $1.97 \pm 0.13\ \mu F\ cm^{-2}$ at $t = 92$ min. *b*, Pretreatment of bladder with colchicine (Col) for 3 h caused, in normal solutions, a $17 \pm 4\%$ increase in C_T. Exposure to half-osmolarity solutions (Hypo) resulted in a rapid increase in capacitance. A return to control solution caused a rapid decrease in C_T to control values (not shown). Actual capacitance was $1.3 \pm 0.13\ \mu F\ cm^{-2}$ (±s.d., $n = 3$) at $t = 0$ (before colchicine treatment). *c*, Bladders were pretreated with cytochalasin B ($35\ \mu M$; CB) for 90 min and then exposed to half-osmolarity solutions (Hypo). Note that there was no change in C_T.

Capacitance, which is directly proportional to real apical membrane area ($1\ \mu F \sim 1\ cm^2$ of membrane)[1] was monitored according to the method of Lewis and de Moura (unpublished), in which the epithelial capacitance is independent of cellular resistances and alterations in cellular or junctional resistances but is directly proportional to apical membrane area. The normal bathing solution had the following composition (mM): 111.2 NaCl, 25 $NaHCO_3$, 5.8 KCl, 2 $CaCl_2$, 1.2 $MgSO_4$, 1.2 K_2HPO_4 and 11.1 glucose, gassed with 95% O_2–5% CO_2 buffered at pH 7.4 at 37 °C.

To investigate the phenomenon of intracellular vesicle translocation and apical membrane fusion, we performed three sets of experiments. In the first experiments, we stretched the epithelium and monitored the capacitance as a function of the stretch. Two approaches can be used to determine vesicle fusion: the first is to correlate tissue dry weight per cm^2 of chamber opening with the membrane area per cm^2 of chamber opening (that is, the capacitance). As the epithelium is stretched, both membrane area per cm^2 (capacitance, C_T) and dry weight per cm^2 will decrease linearly. In the absence of membrane insertion, both C_T ($\mu F\ cm^{-2}$) and dry weight ($mg\ cm^{-2}$) will reach minimum values. If at this point cytoplasmic stores of membrane (having a capacitance of $1\ \mu F\ cm^{-2}$) are inserted into the apical membrane, then C_T will remain constant while dry weight will continue to decrease. Preliminary experiments by Lewis and Diamond[1] demonstrated a biphasic response of capacitance to dry weight. At low levels of stretch both C_T and dry weight (per unit chamber area) decreased in a linear manner. With increased stretch, C_T changed only by 15% (to a minimum of $1.1\ \mu F\ cm^{-2}$) while dry weight decreased by 230%. These data strongly support the proposition that during stretch the urinary bladder epithelial cells change from a goblet shape (having a flat apical membrane) to an oblate spheroid (disk shape) by incorporating cytoplasmic membrane into the apical surface. The capacitance of this added membrane, estimated from the asymptotic capacitance value, is $\sim 1\ \mu F\ cm^{-2}$. Does the cell have to alter its volume during stretch? It is easily demonstrated that given a cell of fixed basolateral area, the apical surface area can be increased by 245% at constant cell volume. Alternatively, one can decrease the weight by 245% at constant apical area.

The second and more satisfactory approach for determining vesicle fusion uses hydrostatic pressure. After mounting the preparation in a chamber modified so that it has no mesh backing, the epithelium is bowed towards the serosal solution by applying an excess of solution to the mucosal solution. At a constant weight (that is, constant mass of exposed tissue) and smooth apical surface area, this epithelium can only increase its capacitance by increasing its surface area. In three separate preparations, excess solution was added to the mucosal chamber. The capacitance (that is, apical membrane area) increased by $18 \pm 3\%$ (±s.e.m., $n = 6$). After eliminating this hydrostatic pressure gradient, the capacitance returned to within $6 \pm 2\%$ (±s.e.m., $n = 6$) of the starting value. This initial experiment indicated that mammalian urinary bladder accommodates stretch by incorporating vesicles from a cytoplasmic pool into the apical membrane and that recovery from stretch results in a removal of membrane back into the cytoplasm. Both the increase and decrease in capacitance occurred within at least 5 min after applying or removing the gradient.

In the second set of experiments, the epithelial cells were induced to swell by replacing the normal mammalian Ringer's solution with one of half osmotic strength. Changes in membrane area were again measured using capacitance as a criterion for area increase[1,5], where an increase in capacitance indicates a proportional increase in apical area. Figure 1*a* illustrates that on application of a hypo-osmotic solution there was no measurable capacitance change until ~15 min later. Apical area then rapidly increased for the next 45 min, reaching a stable value 74% greater than the control level. Returning the solutions to isosmotic resulted in a rapid return of the apical area to near control values. The decrease in capacitance was well defined

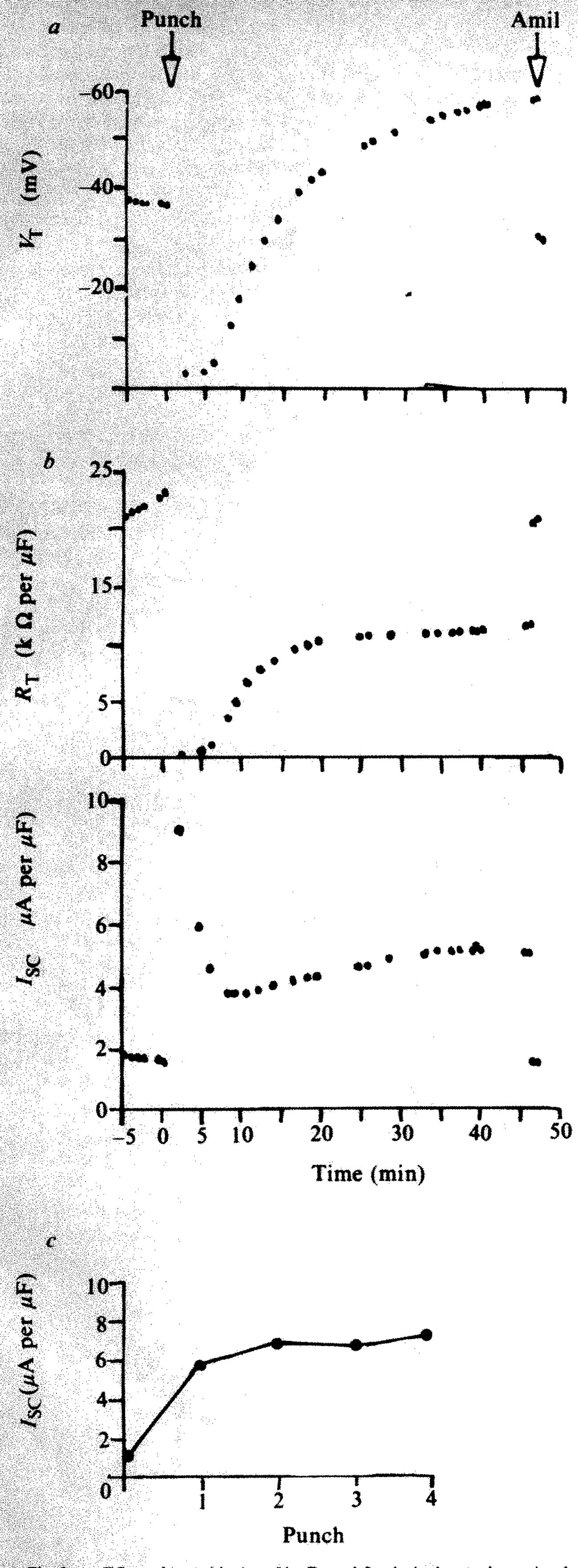


Fig. 2 *a*, Effect of 'punching' on V_T, R_T and I_{sc}. A single punch consisted of rapidly raising and lowering the mucosal solution 10 times (in the absence of a serosal solution). Total time, ~20 s. V_T and R_T were very reduced immediately after a punch. Over a period of 40 min both V_T and R_T increased, as did I_{sc}. V_T and I_{sc} exceeded pre-punch values and, as expected, R_T was reduced. The amiloride-sensitive current (Amil) was much larger than pre-punch values. *b*, Example of the number of punches required to bring the I_{sc} to a stable increased level. The mean value of five experiments was 2.2 ± 0.2 punches.

by a single inverse exponential $C_T(t) = C_T(0)\,e^{-tk}$ where $C_T(t)$ is the normalized capacitance at any time t after restoration of isosmotic solutions, and k is the proportionality constant. In four experiments, k had a value of 0.087 ± 0.015 (±s.d.) min^{-1}, that is, capacitance had decreased to within 37% of control after 11 min.

The cytoplasmic vesicles, apical membrane and tight junctions and desmosomes are all interconnected by a dense microfilament network[6]. The exact role of this network is unknown but has been suggested to aid in mechanical stabilization of the apical membrane during distension, and to be involved in vesicle translocation[1,2]. Disruption of this network should inhibit vesicle translocation during swelling promoted by hypo-osmotic solutions; we tested this by investigating the effects of colchicine[7] (a microtubule blocking agent) and cytochalasin B[8] (CB, a microfilament disrupting agent). Pre-incubation for 180 min with colchicine (0.1 mM) resulted in an increase ($17 \pm 4\%$) in capacitance, and a much more rapid onset of the capacitance change after exposure to hypo-osmotic solutions (Fig. 1*b*). The decrease after returning to the normal Ringer's followed the same time course as untreated bladders. Pre-incubation with cytochalasin B (~35 μM for 60 min) produced no change in the epithelial capacitance but it did inhibit the capacitance increase due to hypo-osmotic serosal solutions (Fig. 1*c*). This therefore indicates that increases in apical membrane area are dependent on intact microfilaments.

To determine whether the decrease in apical area after returning the epithelium to isosmotic Ringer's was also filament-dependent, the epithelium was first incubated for 60 min in hypo-osmotic Ringer's and then for a further 60 min in the presence of CB. The epithelium was then returned to the isosmotic Ringer's (in the absence of CB). As in the control studies, the capacitance decreased as a single exponential. However, the proportionality constant decreased to $0.046 \pm 0.002\ \text{min}^{-1}$ compared with $0.087 \pm 0.015\ \text{min}^{-1}$, thus the time to reach 37% of control was increased from 11 to 22 min. These data indicate that the microfilaments are necessary not only for exocytosis of vesicles but also for endocytosis. Microtubule-blocking agents seem to cause a partial disruption of the cell volume regulating system as capacitance changes were noted in the absence of hypo-osmotic solutions, and on exposure to a hypo-osmotic solution, capacitance increased more rapidly than for controls.

What is the role of the vesicles in electrolyte transport? Lewis and Wills[4] reported that the Na^+–K^+ selective permeability (P_{Na}/P_K) of the amiloride pathways is 2.3:1 in control animals and increases to 9:1 in animals maintained on low Na^+ diets (that is, there is an increase in endogenous aldosterone level). In the following experiment we determined the selectiviy of the newly inserted membrane in diet and control animals. We used a method involving mechanical perturbation of the epithelium, which maximized the insertion of vesicles into the apical membrane. Briefly, this method (termed 'punching') involved removing the solution from both mucosal and serosal chambers. Next, we rapidly filled and emptied the mucosal chamber of solution–this compresses the epithelium against a nylon mesh. Finally, we filled both chambers with normal bathing solution.

Figure 2 illustrates the effect of punching on transepithelial potential (V_T, referenced to the serosal solution), transepithelial resistance (R_T, normalized to capacitance) and short-circuit current (I_{sc}, calculated as V_T/R_T). V_T and I_{sc} were greater and R_T less than the respective values before punching. A significant finding was that punching caused no change in apical capacitance but significantly enhanced (from 0.4 ± 0.1 to 4.7 ± 0.7 μA per μF; ±s.e.m., $n = 10$) amiloride-sensitive Na^+ transport. The Na^+–K^+ selective permeability (P_{Na}/P_K), calculated using the method of Lewis and Willis[4], was greatly increased (30:1) compared with the pre-punch value but was not significantly different between control and diet animals. The absolute apical membrane Na^+ permeability was, however, increased in diet compared with control animals. A possible explanation for the

difference in P_{Na}/P_K between pre-punched and punched preparations might be related to an as yet unknown property of urine (for example, proteolytic enzymes) to cause degradation of the selectivity properties of the amiloride pathways. How does punching cause an increase in Na^+ transport but not capacitance? We hypothesize that rapid filling and emptying of the mucosal chamber drives the apical membrane and cytoplasmic vesicles towards each other. On contact, the apical membrane may capture some vesicles as they 'kiss' onto the apical membrane. After the pressure pulse, the apical membrane rebounds, and because the vesicles are tethered by microfilaments, this results in a removal of some vesicles from the apical membrane. Evidence which partly supports such a mechanism is that pretreatment of the epithelium with CB followed by punching, does not elicit an increase in Na^+ transport.

Because these vesicles contain amiloride-sensitive Na^+ transport systems, we have devised a method for estimating the area of the apical membrane relative to that of the cytoplasmic pool. This method is based on the assumption that all the apical membrane and cytoplasmic vesicles are readily exchangeable. First, we 'punched' the epithelium to a constant (elevated) Na^+ transport rate, then we eliminated irreversibly all apical Na^+ pathways using the sulphydryl-reactive agent *p*-chloromercuribenzene sulfonic acid (PCMBS). Then we punched the epithelium to a new (lower) Na^+ transport rate. By comparing the Na^+ transport rates before and after PCMBS treatment and punching, we were able to estimate the area ratios. The ratio of pool/apical area was 3.3 ± 1.2, in good agreement with morphological studies[2].

This represents the first report of a physiological role for the cytoplasmic vesicles in normal functioning of the mammalian urinary bladder. These vesicles serve a dual role, first by allowing an increase in the storage volume of the urinary bladder without a concomitant alteration in the cellular volume. Thus, although the cells undergo dramatic geometric changes, there is a minimal alteration in intracellular ionic activities. Insertion and removal of the vesicles depends on an intact microfilament network.

Second and most fascinating, the vesicles contain Na^+ transport pathways whose density is increased by aldosterone. This phenomenon raises the possibility of membrane transport regulation by insertion and removal of vesicles containing the transport protein.

We thank Dr N. K. Wills for critical review of this manuscript and W. Alles for technical assistance. This work was supported by NIH grant AM 20851.

Received 21 December 1981; accepted 1 April 1982.

1. Lewis, S. A. & Diamond, J. M. *J. Membrane Biol.* **28**, 1–40 (1976).
2. Minsky, B. D. & Chlapowski, F. J. *J. Cell Biol.* **77**, 685–697 (1978).
3. Lewis, S. A. & Diamond, J. M. *Nature* **253**, 747–748 (1975).
4. Lewis, S. A. & Wills, N. K. *Ann. N. Y. Acad. Sci.* **372**, 56–63 (1981).
5. Clausen, C., Lewis, S. A. & Diamond, J. M. *Biophys. J.* **26**, 291–318 (1979).
6. Porter, K. R., Kenyon, K. & Badenhausen, S. *Protoplasma* **63**, 262–274 (1967).
7. Soifer, D. (ed.) *Ann. N. Y. Acad. Sci.* **253** (1975).
8. Tanenbaum, S. W. (ed.) in *Cytochalasins: Biochemical and Cell Biological Aspects* (North-Holland, Amsterdam, 1978).

Na^+ transport and flux ratio through apical Na^+ channels in toad bladder

Lawrence G. Palmer

Department of Physiology, Cornell University Medical College, 1300 York Avenue, New York, New York 10021, USA

In many epithelia, the rate of NaCl reabsorption is determined by the activity of a Na^+ transport system in the outer (apical) membrane. In frog skin and toad bladder, this system is thought to involve transmembrane pores through which Na^+ ions move down an electrochemical activity gradient[1–5]. It has been shown, however, that increased intracellular Na^+ concentration leads to reduction in unidirectional Na^+ influx[6,7], raising the possibility that ions do not move through the channel independently. The Ussing flux ratio equation is one test for independence of passive ion movements[8]. In particular, single-filing of ions through long pores can account for effects such as reduction of unidirectional fluxes by ions on the opposite (*trans*) side of the membrane; this type of transport mechanism is characterized by flux ratio exponents >1 (refs 9–11). I have evaluated the flux ratio exponent (n') for the apical Na^+ channel in the toad bladder as the ratio of tracer permeability to electrical conductance at electrochemical equilibrium[9–11], that is, $n' = RT/F^2(G_{Na}/J_{Na})$, where J_{Na} and G_{Na} are, respectively, unidirectional flux and the conductance for Na^+ ions through the channel in the absence of net Na^+ movement. I report here values of n' of 1.15 ± 0.10 and 1.08 ± 0.07 for activities of 40 mM and 10 mM Na^+, respectively, in the outer solution. This channel is therefore occupied by at most one Na^+ ion at these activities.

Urinary bladders from toads (*Bufo marinus*) were mounted in Lucite chambers and short-circuited as described previously[12]. The serosal surface was bathed in KCl–sucrose medium to depolarize the basolateral membranes. In these conditions, the resistance and potential difference across the basolateral membranes fall sharply[12], and transepithelial impedance measurements have shown that almost all the transcellular resistance is located at the apical membrane[13]. Analysis of voltage transients showed that this condition is maintained in the Na^+-loaded state in the presence of ouabain (my unpublished results). Conductance was measured by displacing the transepithelial voltage by 10 mV and G_{Na} was calculated as the difference in transepithelial conductance obtained by replacing mucosal Na^+ with K^+ or, equivalently, by adding 10^{-4} M amiloride to the mucosal solution[12]. In the Na^+-loaded state, the amiloride-sensitive current–voltage (I–V) relationship becomes linear near $V = 0$ (ref. 12), so that G_{Na} is a reliable measure of conductance. All conductance measurements were corrected for solution resistance. The intracellular Na^+ ion activity $[Na^+]_i$ was estimated from the amiloride-sensitive short-circuit current (I_{Na}) and G_{Na}, using the constant-field equation. This is a rough estimate, and does not affect the other measured or calculated parameters.

J_{Na} was measured as the uptake of ^{22}Na into the cell from the mucosal solution during brief (15–20 s) exposures to the isotope, following the protocol of Biber and Curran[14]. Uptake was corrected for extracellular Na^+ using ^3H-mannitol as a marker. I also made a small correction for amiloride-insensitive uptake (J_{Na}, corrected for extracellular Na^+, in the presence of 10^{-4} M amiloride). In control conditions, uptake was found to be linear with time for at least 24 s, and J_{Na} was almost equal to I_{Na}, in agreement with previous reports[7,14,15] (Fig. 1).

To achieve electrochemical equilibrium for Na^+ across the membrane, ouabain was added to the serosal medium to block Na^+ extrusion from the cell. As shown in Fig. 2, I_{Na} declined rapidly at first, and then more slowly. $[Na^+]_i$ rose from initial values of <5 mM to ~40 mM, and then remained fairly constant. G_{Na} increased initially, as expected for an increase in $[Na^+]_i$, and then fell, indicating a decline in Na^+ permeability.

When $[Na^+]_i$ reached 40 mM, the mucosal solution containing 80 mM NaCl was replaced with one of 40 mM NaCl. Instan-

Table 1 Flux-ratio exponent (n') of apical membrane Na^+ channel

Mucosal Na^+ concentration (mM)	J_{Na} (pEq s^{-1})	G_{Na} (mS)	n' $[RT/F^2(G_{Na}/J_{Na})]$
10 ($n = 14$)	233 ± 49	0.92 ± 0.19	1.08 ± 0.07
40 ($n = 14$)	278 ± 67	1.09 ± 0.19	1.15 ± 0.10

Values represent mean ± s.e.m.

taneously, I_{Na} declined to near zero, and G_{Na} fell slightly, whereas $[Na^+]_i$ did not change. The residual I_{Na}, slightly negative in Fig. 2, was not significantly different from zero in either series of experiments. The absence of a Na^+ current in the presence of a Na^+ conductance demonstrated that electrochemical equilibrium had been achieved. At this time, G_{Na} and J_{Na} were measured, and n' was computed. In further experiments, a similar protocol was followed, except that shorter Na^+ loading periods were used to increase $[Na^+]_i$ to ~10 mM, and G_{Na} and J_{Na} were measured with 10 mM Na in the mucosal solution. Mean computed values of n' were 1.15 at 40 mM Na^+ and 1.08 at 10 mM Na^+ (Table 1); neither value is significantly different from 1 ($P>0.1$).

The results indicate that single-filing of ions does not occur to any appreciable extent at these Na^+ concentrations; the channel is occupied by at the most one Na^+ ion at a time. Single-filing therefore does not account for the decline in Na^+ permeability resulting from increased intracellular Na^+ observed here (Fig. 2) and by others[7,8,16]. The slow fall in G_{Na} at constant $[Na^+]_i$ implies that the decline in permeability may be an indirect result of high intracellular Na^+, perhaps mediated by changes in intracellular Ca^{2+} (refs 7, 17, 18).

Furthermore, the data indicate that the saturation of Na^+ influx with increasing mucosal Na^+ (refs 5, 14, 19) does not reflect carrier or shuttle-type kinetics, as these mechanisms, in general, show positive coupling between unidirectional fluxes, and flux-ratio exponents <1 at concentrations at or above the dissociation constant for the carrier site. The results are consistent with those of Fuchs *et al.*[5], Van Driessche and Lindemann[2], and Aceves and Cuthbert[20] who found that saturation was accounted for by a decreased number of conducting channels, and not by saturation of single transporting sites.

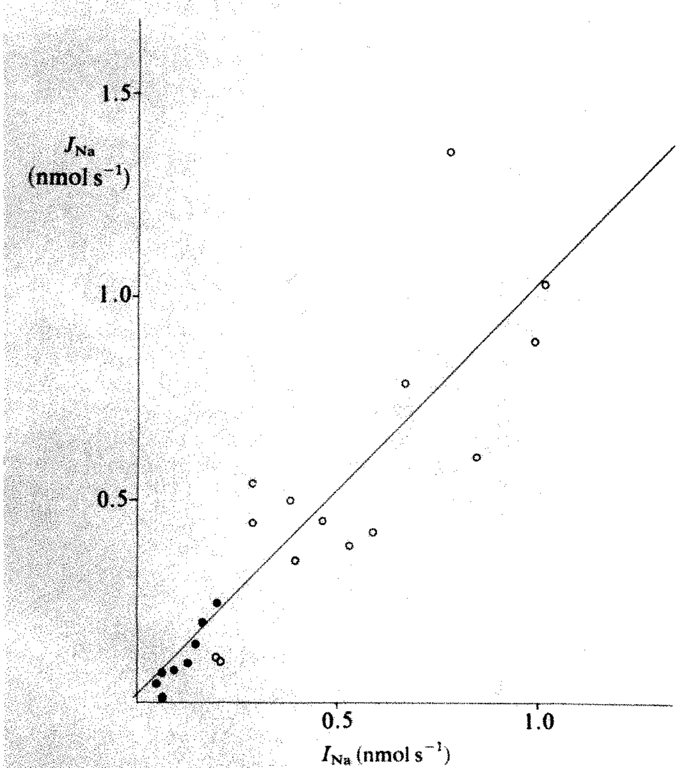


Fig. 1 Correlation of influx of Na^+ with net flux. Hemibladders were incubated with KCl–sucrose medium (composition (in mM): 85 KCl, 50 sucrose, 1 $CaCl_2$, 0.5 $MgSO_4$, 5 glucose, 3.5 K-phosphate, pH 7.7) on the serosal side and either 80 mM Na^+ concentration (composition (in mM): 108 NaCl, 7 KCl, 1 $CaCl_2$, 0.5 $MgSO_4$, 3.5 K-phosphate, pH 6.0) (○) or 10 mM Na^+ (replaced by choline) (●) on the mucosal side. Na^+ permeability was optimized at pH 6. Membrane area was 3 cm^2. For measurements of J_{Na}, the mucosal solution was replaced with labelled solution of identical composition except for the addition of ^{22}Na (5 μCi ml^{-1}, carrier-free), ^{3}H-mannitol (10 μCi ml^{-1}, 17 Ci mmol^{-1}) and 5 mM cold mannitol. The straight line represents a least-squares linear regression: $J_{Na}=1.01\ I_{Na}+0.018$ ($r=0.88$).

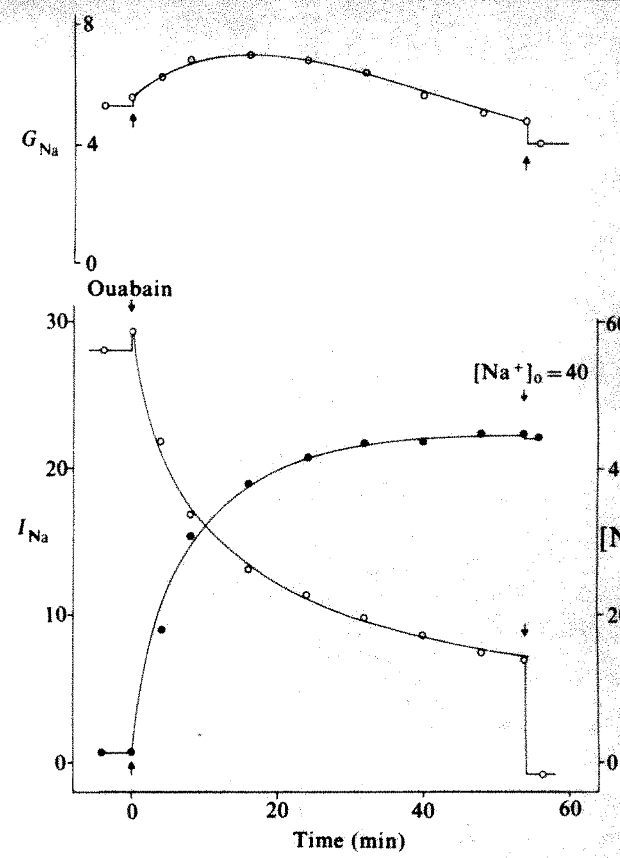


Fig. 2 Protocol for Na^+-loading epithelial cells. Hemibladders were preincubated with KCl–sucrose medium on the serosal side, and 80 mM Na^+ on the mucosal side. Membrane area was 3 cm^3. At time 0, ouabain (5 mM) was added to the serosal medium. G_{Na} (mS) and I_{Na} (μA) were calculated as the amiloride-sensitive conductance and short-circuit current, respectively. $[Na^+]_i$ (mM) was estimated from the constant-field equation assuming no membrane potential using: $G_{Na}=(F^2/RT)P_{Na}([Na^+]_o+[Na^+]_i)/2$ and $I_{Na}=(F)P_{Na}([Na^+]_o-[Na^+]_i)$ where P_{Na} is the Na^+ permeability and $[Na^+]_o$ and $[Na^+]_i$ are the Na^+ activities in the mucosal solution and in the cell, respectively. After 55 min, the Na^+ activity in the mucosal solution was rapidly decreased to 40 mM by replacement with K^+. G_{Na} and J_{Na} were determined immediately afterwards, during the period in which $I_{Na}=0$.

Finally, deviations in transepithelial flux ratios from those predicted for passive, independent ion movement have been used to study active Na^+ transport in tight epithelia[21]. The results reported here imply that transport of Na^+ across the apical membrane Na^+ channel satisfies this condition for independence, and that the reported deviations of transepithelial flux ratio exponents from unity reflect the activity of the serosal Na^+ pump.

I thank N. Speez for technical assistance. Financial support was provided by the Whittaker Foundation and USPHS grant AM 27847.

Received 21 January; accepted 13 April 1982.

1. Lindemann, B. & Van Driessche, W. *Science* **195**, 292–294 (1977).
2. Van Driessche, W. & Lindemann, B. *Nature* **282**, 519–520 (1979).
3. Li, J. H.-Y., Palmer, L. G., Edelman, I. S. & Lindemann, B. *J. Membrane Biol.* **64**, 77–90 (1982).
4. Palmer, L. G., Li, J. H.-Y., Lindemann, B. & Edelman, I. S. *J. Membrane Biol.* **64**, 91–102 (1982).
5. Fuchs, W., Hviid Larsen, E. & Lindemann, B. *J. Physiol., Lond.* **267**, 137–166 (1977).
6. Erlij, D. & Smith, D. *J. Physiol., Lond.* **228**, 221–239 (1973).
7. Chase, H. S. Jr & Al-Awqati, Q. *J. gen. Physiol.* **77**, 693–712 (1981).
8. Ussing, H. H. *Acta physiol. scand.* **19**, 43–56 (1949).
9. Hodgkin, A. L. & Keynes, R. D. *J. Physiol., Lond.* **128**, 61–88 (1955).
10. Heckmann, K. *Biomembranes* **3**, 127–153 (1972).
11. Hille, B. & Schwartz, W. *J. gen. Physiol.* **72**, 409–442 (1978).
12. Palmer, L. G., Edelman, I. S. & Lindemann, B. *J. Membrane Biol.* **57**, 59–71 (1980).
13. Warncke, J. & Lindemann, B. *Adv. physiol. Sci.* **3**, 129–133 (1980).
14. Biber, T. U. L. & Curran, P. F. *J. gen. Physiol.* **56**, 83–99 (1970).

15. Thompson, S. M. & Dawson, D. C. J. Membrane Biol. 42, 357–374 (1978).
16. MacRobbie, E. A. C. & Ussing, H. H. Acta physiol. scand. 53, 348–365 (1961).
17. Grinstein, S. & Erlij, D. Proc. R. Soc. B202, 353–360 (1978).
18. Taylor, A. & Windhager, E. E. Am. J. Physiol. 236, F505–F512 (1979).
19. Frazier, H. S., Dempsey, E. F. & Leaf, A. J. gen. Physiol. 45, 529–543 (1962).
20. Aceves, J. & Cuthbert, A. W. J. Physiol., Lond. 295, 491–504 (1979).
21. Ussing, H. H., Eskesen, K. & Lim, J. in Epithelial Ion and Water Transport (eds MacKnight, A. D. C. & Leader, J. P.) 257–264 (Raven, New York, 1981).

Inhibitors of polyamine biosynthesis block human cytomegalovirus replication

A. S. Tyms & J. D. Williamson

Department of Virology, St Mary's Hospital Medical School, Paddington, London W21PG, UK

The antiviral activities of the few compounds available that are effective in the clinical treatment of virus infections are directed against nucleic acid synthesis[1,2]. Attention has been drawn recently to the need to develop new strategies, based on specific inhibitors of other biopolymers with essential functions in virus replication[2]. Polyamines may prove suitable targets, as these aliphatic bases are found in all animal cells, where they seem to have essential roles in the synthesis and stabilization of macromolecules[3–5]. Here we describe how human diploid fibroblasts (MRC5 cells) infected with human cytomegalovirus (CMV) have elevated levels of polyamines, and the polyamine antimetabolite methylglyoxal bis(guanylhydrazone) (MGB G) has potent antiviral activity. The growth of human CMV is also blocked by α-difluoromethylornithine (DFMO), a highly specific inhibitor of the initiation of polyamine biosynthesis. Cytochemical studies show this inhibitory effect is directed against early events in the virus replication cycle.

Infection of human diploid fibroblasts with human CMV is known to stimulate ornithine decarboxylase (EC 4.1.1.17) activity[6]. This enzyme catalyses the first step of polyamine biosynthesis, the formation of putrescine[3]. Another study has shown increased levels of [1,4-^{14}C]putrescine uptake into MRC-5 cells infected with human CMV[7]. However, although radiolabelled polyamines (spermidine and spermine) were recovered at times from 24 h post-infection (p.i.), specific activities were not determined. We have now quantitatively analysed polyamine biosynthesis in MRC-5 cells infected at high multiplicity with human CMV strain AD169. The infected cells were recovered at 72 h p.i. and acid-soluble fractions were assayed for polyamines by the fluorometric measurement of their dansylated derivatives, after separation by TLC. The concentrations of spermidine and spermine in infected cells were 1.5- and 12.4-fold higher, respectively, than in uninfected control cells (Table 1). The higher specific activities of both polyamines in infected cells are consistent with the increased utilization of radiolabelled putrescine described in the earlier study[7], showing clearly that infection with human CMV stimulates polyamine biosynthesis. The requirement for continued polyamine metabolism during human CMV replication is reflected in the antiviral effect of specific inhibitors of polyamine biosynthesis. Both MGBG, an inhibitor of spermidine and spermine biosynthesis, and α-methylornithine, a competitive inhibitor of ornithine decarboxylase, have been shown to prevent the production of infectious virus[8]. These results have encouraged further *in vitro* studies of the antiviral properties of MGBG in comparison with the nucleoside analogues 9-(2-hydroxyethoxymethyl) guanine (acyclovir) and 5-iodo-2′-deoxyuridine (IUdR), two drugs that have been used clinically for the treatment of human CMV infections[1,9]. The inhibitory effect of each compound against eight different virus isolates was determined by plaque reduction methods (Table 2) and the median values (ED_{50}) for the antiviral effect of acyclovir (33.5 μM), IUdR (5.3 μM) and MGBG (0.6 μM) demonstrate

Table 1 Spermidine and spermine content of uninfected and human CMV-infected MRC-5 cells

Cells	Polyamine	nmol per 2×10^{-7} cells	d.p.m. $\times10^{-4}$ per 2×10^{7} cells	Specific activity (mCi mmol^{-1})
Uninfected	Spermidine	75.8	4.11	0.25
	Spermine	18.5	0.82	0.19
Infected	Spermidine	110.0	16.15	0.66
	Spermine	230.0	42.34	0.83

MRC-5 cells were grown and maintained as described previously[7]. Confluent monolayers were mock-infected or infected with human CMV (strain AD169) using five plaque-forming units (PFU) per cell. All virus stocks were prepared from extracellular virus as described elsewhere[16]. At 24 h p.i. [1,4-^{14}C]putrescine (116 mCi mmol^{-1}) was added to the maintenance medium to a final concentration of 0.1 μCi ml^{-1} and the cells were recovered 48 h later. Polyamines were extracted and separated as their dansyl derivatives by TLC as described previously[7]. After extraction of the dansylated polyamines in dioxan, fluorescence relative to known amounts of appropriate standards were measured using a Locarte fluorimeter. Duplicate determinations were made and representative results are presented. The amount of radioactivity present in each extract was measured in a Packard liquid scintillation spectrometer.

the greater potency of the polyamine inhibitor. This effect on virus growth, which is readily reversed by exogenous spermidine (unpublished results), correlates with the reduction in polyamine biosynthesis in CMV-infected cells treated with MGBG[8]. However, this inhibitor is reported to have other effects on cellular functions which may be unrelated to polyamine metabolism[11].

DFMO is a highly specific inhibitor of polyamine metabolism with remarkably low toxicity[12,13]. The compound is catalytically activated by ornithine decarboxylase, which it irreversibly inhibits[14]. The antiviral properties of DFMO were investigated by measuring its effect on the production of infective progeny virus. MRC-5 cells infected with human CMV were maintained in medium containing various concentrations of the compound, and infectivity titres were measured at 120 h p.i., a time after the exponential phase in the virus growth cycle[7]. There was a marked reduction in the amount of infective virus produced in cultures treated with 10 mM DFMO (see Fig. 1 legend). Lower concentrations had less effect on virus replication, and subsequent removal of the inhibitor from infected cultures restored virus growth. The replication of human CMV is characterized by the formation of intranuclear inclusions which are the sites of virus-specific DNA synthesis[15,16]. These inclusions, visualized by acridine orange staining, develop progressively from 24 h p.i. to reach maximum size at 120 h p.i. (Fig. 1*b,c*). The production is inhibited completely by 10 mM DFMO (Fig. 1*f*), but with lower concentrations of the inhibitor the increase in virus yield is accompanied by development of the cytomegalic effect (Fig. 1*d,e*). These cytochemical studies show that the antiviral properties of DFMO are directed against early events in the CMV replication cycle associated with the synthesis or utilization of virus-specific DNA. Such results are consistent with the antiproliferative effects of DFMO described in other systems[12].

Table 2 A comparison of the antiviral effect (ED_{50}) of MGBG and two nucleoside analogues against eight strains of human CMV

Compound	AD169	Kerr	Rawles	BL	BR	LA	RE	WE
Acyclovir (μM)	78	15	28	39	10	52	25	68
IUdR (μM)	4.2	5.0	3.3	7.7	8.8	9.9	5.5	4.5
MGBG (μM)	0.9	0.8	1.8	0.5	0.6	0.5	0.5	0.5

(Columns AD169 to WE are headed "Virus strains".)

Three laboratory strains of CMV (AD169, Kerr, Rawles) and five clinical isolates with a low passage history were included in this study. In each experiment, dose response curves were constructed for acyclovir, IUdR and MGBG against the different strains of human CMV. A plaque reduction method was used to determine ED_{50} values in a manner similar to that described by Collins and Bauer[10]. In brief, confluent MRC-5 cultures infected with about 100 PFU per culture were subsequently maintained in medium containing 0.5% agarose and supplemented with doubling concentrations of each compound. After 7–10 days incubation at 36°C, infected cultures were fixed in 5% formalin in phosphate-buffered saline and plaques were counted after staining with methylene blue. The ED_{50} values were determined from plots of the percentage plaque reduction against the $\log_{10}$ molar concentration of each compound.

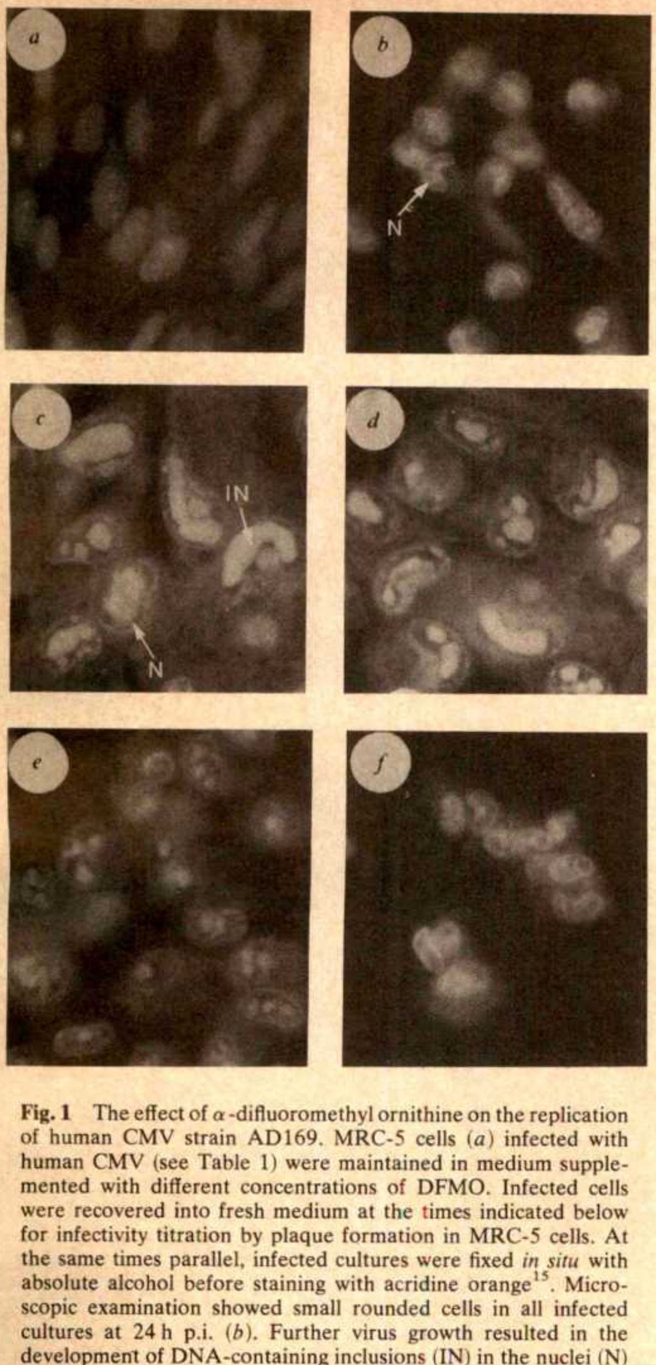


Fig. 1 The effect of α-difluoromethyl ornithine on the replication of human CMV strain AD169. MRC-5 cells (*a*) infected with human CMV (see Table 1) were maintained in medium supplemented with different concentrations of DFMO. Infected cells were recovered into fresh medium at the times indicated below for infectivity titration by plaque formation in MRC-5 cells. At the same times parallel, infected cultures were fixed *in situ* with absolute alcohol before staining with acridine orange[15]. Microscopic examination showed small rounded cells in all infected cultures at 24 h p.i. (*b*). Further virus growth resulted in the development of DNA-containing inclusions (IN) in the nuclei (N) of infected cells and, in addition, there was a marked increase in cell size (cytomegalic effect) by 120 h p.i. (*c*). In the presence of 10 mM DFMO, virus growth was inhibited and infected cells remained in the early, rounded form (*f*) but production of infective virus together with development of the intranuclear inclusions was seen with lower concentrations of the inhibitor (*d*, 2.5 mM; and *e*, 5 mM). Time p.i. (hours); *b*, 24; *c*, *d*, *e* and *f*, 120. Virus titre (PFU per 10^6 cells); *b*, $<10^2$; *c*, 2.2×10^6; *d*, 6.4×10^6; *e*, 7.5×10^5; *f*, 6.5×10^2.

Further development of these observations by *in vivo* studies is limited by the lack of experimental animal models of the human disease. Polyamine inhibitors, however, have been used clinically as chemotherapeutic agents. MGBG has been used in the treatment of acute leukaemia[17] with considerable benefit to the patient if the drug is monitored appropriately[18]. Clinical prospects are also good for the use of DFMO as an antiprotozoal agent[19] and for the treatment of hyperproliferative states[20]. In each instance the efficacy of chemotherapy may be related to inhibition of polyamine metabolism[3,19,21]. Human CMV is the most important known infectious cause of mental retardation and a significant complication in immunocompromised hosts, particularly allograft recipients[22]. No effective treatment of infections with this virus is currently available and there are certain limitations to prophylactic measures based on virus vaccines[23]. Our observations show that polyamine metabolism is essential for the replication of human CMV and polyamine inhibitors may be suitable drugs for the treatment of CMV infections.

DFMO and acyclovir were gifts from Centre de Recherche Merrell International and Wellcome Research Laboratories, respectively. We acknowledge the financial support of the Joint Standing Research Committee of St Mary's Hospital and an anonymous donor.

Received 8 March; accepted 21 April 1982.

1. Bauer, D. J. *The Specific Treatment of Virus Disease* (MTP, Lancaster, 1977).
2. Cohen, S. S. *Science* **205,** 964–971 (1979).
3. Janne, J., Pösö, H. & Raina, A. *Biochim. biophys. Acta* **473,** 241–293 (1978).
4. Canellakis, E. S., Viceps-Madore, D., Kyriakidis, D. A. & Heller, J. S. *Curr. Topics cell. Regulation* **15,** 155–202 (1979).
5. Cohen, S. S. *Nature* **274,** 209–210 (1978).
6. Isom, H. C. *J. gen. Virol.* **42,** 265–278 (1979).
7. Tyms, A. S. & Williamson, J. D. *J. gen. Virol.* **48,** 183–191 (1980).
8. Tyms, A. S., Scamans, E. & Williamson, J. D. *Biochem. biophys. Res. Commun.* **86,** 312–318 (1979).
9. Plotkin, S. A. *et al.* in *The Human Herpesviruses* (eds Nahmias, A. J., Dowdle, W. R. & Schinazi, R. F.) 411 (Elsevier, New York, 1981).
10. Collins, P. & Bauer, D. J. *Ann. N.Y. Acad. Sci.* **284,** 49–59 (1977).
11. Pathak, S. N., Porter, C. W. & Dave, C. *Cancer Res.* **37,** 2246–2250 (1977).
12. Mamont, P. S., Duchesne, M-C, Grove, J. & Bey, P. *Biochem. biophys. Res. Commun.* **81,** 58–66 (1978).
13. Sjoerdsma, A. *Clin. pharm. Ther.* **30,** 3–22 (1981).
14. Metcalf, B. W. *et al. J. Am. chem. Soc.* **100,** 2551–2553 (1978).
15. McAllister, R. M., Straw, R. M., Filbert, J. E. & Goodheart, C. R. *Virology* **19,** 521–531 (1963).
16. Huang, E.-S., Chen, S.-T. & Pagano, J. S. *J. Virol.* **12,** 1473–1481 (1973).
17. Levin, R. H., Henderson, E., Karon, M. & Freireich, E. J. *Clin. Pharmac. Ther.* **6,** 31–42 (1964).
18. Siimes, M., Seppänen, P., Alhonen-Hongisto, L. & Jänne, J. *Int. J. Cancer* **28,** 567–570 (1981).
19. Bacchi, C. J., Nathan, H. C., Hutner, S. H., McCann, P. P. & Sjoerdsma, A. *Science* **210,** 332–334 (1980).
20. Grosshans, E. *et al. Ann. Dermatol. Vener.* **107,** 377–387 (1980).
21. Böhlen, P. *et al. Eur. J. clin. Invest.* **8,** 215–218 (1978).
22. Timbury, M. & Emond, E. *J. clin. Path.* **32,** 859–881 (1979).
23. Lang, D. *Rev. infect. Dis.* **2,** 449–458 (1980).

Derivation of macrophage-like lines from the pre-B lymphoma ABLS 8.1 using 5-azacytidine

A. W. Boyd & J. W. Schrader

The Walter and Eliza Hall Institute of Medical Research, Post Office Royal Melbourne Hospital, Victoria 3050, Australia

Variation in the degree of methylation of DNA seems to be one mode of regulating gene expression in eukaryotic cells[1–7]. The relationship between DNA demethylation and gene activation observed in globin[4–6] and viral[2] genes, together with evidence that alterations in the degree of DNA methylation of a gene are heritable[8], although not with 100% fidelity[9], have suggested that this may be a mechanism of control of differentiation. Futhermore, exposure to the demethylating drug 5-azacytidine (5-AC) causes differentiation of 3T3 cells into striated muscle cells, chondrocytes and adipocytes[7]. Subsequent studies have shown that these effects are due to DNA demethylation[10]. In view of these observations, we have now attempted to modify several continuous B-cell lines with 5-AC. Following exposure of the pre-B lymphoma ABLS 8.1 to 5-AC, we have derived cloned cell lines which possess macrophage-like characteristics not expressed by ABLS 8.1. Similar macrophage-like cell lines were obtained in two independent experiments; they have been re-cloned and remain stable after 4 months of continuous culture.

ABLS 8.1 is a pre-B-cell tumour derived from a BALB/c mouse infected with Abelson virus[11]. The line was originally

classified as B cell because of the presence of surface immunoglobulin[12]. We can detect neither surface nor cytoplasmic immunoglobulin in these cells, but recently the classification of ABLS 8.1 was confirmed by demonstrating that rearrangement of an immunoglobulin variable region gene has occurred at both immunoglobulin heavy chain loci and at one κ chain gene locus (ref. 13 and S. Gerondakis, O. Bernard and E. Webb, in preparation), processes that seem unique to B-cell differentiation.

Cloned ABLS 8.1 cells were cultured at 2.5×10^5 cells per ml in the presence of various concentrations of 5-AC. Of cells cultured with 1 μM 5-AC and fed after 3 days with fresh medium again containing 1 μM 5-AC, only 20% of the initial number were present after 7 days. This population was cloned at limit dilution in liquid culture medium [Dulbecco's modified Eagle's medium (DME), 10% fetal calf serum, 5×10^{-5} M 2-mercaptoethanol (2-ME)]. Of 192 wells, 16 showed evidence of cell growth and these cells were expanded further and examined. Nine of the lines closely resembled ABLS 8.1 and on subsequent testing were indistinguishable from the parent cell line. The remaining seven cell lines were larger (mean cell volume 550 compared with 500 μm^3 for ABLS 8.1) and were characteristically pleomorphic; ~5–20% of the cells had an elongate amoeboid or a stellate appearance and 1–2% adhered to the culture dishes. In contrast, ABLS 8.1 grew as a uniform population of round, non-adherent cells. Furthermore, unlike ABLS 8.1, the lines were able to grow in medium lacking 2-ME. All the lines were re-cloned in the absence of 2-ME in both agar gel culture and liquid culture; the morphological characteristics and ability to grow in the absence of 2-ME were heritable and occurred in all the clones.

One obvious possibility was that a contaminating cell type had been present in the initial cell population of ABLS 8.1 cells. This was excluded in three ways. First, the only tumours carried in tissue culture at the Hall Institute having characteristics resembling the lines putatively derived from ABLS 8.1 were the macrophage-like lines WEHI 3B and WEHI 265. Both, however, exhibited morphological differences from the 5-AC-treated cells and furthermore possessed characteristic metacentric chromosomes. ABLS 8.1 has no detectable chromosomal markers and is diploid. Chromosome spreads prepared [14] from ABLS 8.1, WEHI 3B, WEHI 265 and the newly derived lines showed that ABLS 8.1, WEHI 3B and WEHI 265 had the expected karyotypes. None of the new cell lines had metacentric chromosomes and all had 40 chromosomes, which appeared identical to those of ABLS 8.1. The second way to exclude the possibility of contamination was to repeat the procedure with a new batch of ABLS 8.1 cells. Cloned ABLS 8.1 cells were cultured in agar with 5-AC as described above. To control for the possibility that 5-AC was not essential for the selection of a hardier contaminant of variant line not requiring 2-ME, a parallel culture of 10^7 of the same clone of ABLS 8.1 cells was established in medium lacking both 5-AC and 2-ME. All these cells appeared dead after 3 days and no viable cells were detected over the ensuing 4 weeks of culture. Cells from the second batch of 5-AC-treated cultures

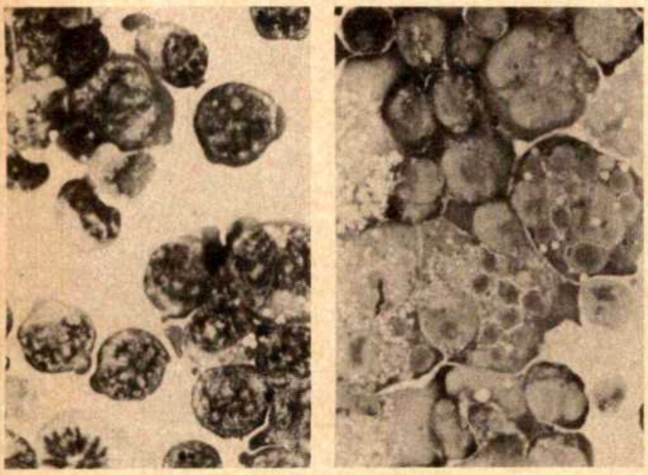


Fig. 1 5-AC-treated ABLS 8.1 cells are morphologically different from untreated ABLS 8.1 cells. Left, cytocentrifuge smear of ABLS 8.1 cells (May–Grünwald–Giemsa). Right, smear of a cloned line of cells derived from 5-AC-treated ABLS 8.1 cells. Three large vacuolated cells which have phagocytosed immunoglobulin-coated sheep red cells are shown.

Table 1 Comparison of the properties of ABLS 8.1 cells and lines derived using 5-AC

	ABLS 8.1	Cell line from first set of 5-AC-treated cultures	Cell line from second set of 5-AC-treated cultures
2-ME dependence	+	–	–
Phagocytosis:			
Latex*	1±0.5	22±3	15±4
Ig-coated SRBCs†	0	12±2	7±3
Fc receptor‡	2±1	63±7	56±8
Membrane Ig§	—	—	—
Membrane Thy-1§	—	—	—
Secreted Ig‖	—	—	—
MAC 1¶	0	6±1.4	3±1.1
Esterase #	±	++	++
Factor production:			
GM-CSF	—	—	—
TCGF	—	—	—
PSF	—	—	—
Presence of Ia antigen	—	—	—
Natural killer activity	—	—	—
Response to macrophage-activating factor**	—	—	—

GM-CSF, granulocyte–macrophage colony-stimulating activity, assayed by colony formation in agar culture[20]. TCGF, T-cell growth factor activity, assayed as previously described[21]. PSF, P-cell factor which promotes growth of a mast cell progenitor; assayed as previously described[22]. Ia antigen was detected by incubating with a monoclonal antibody[23] at 4°C for 1 h, washing and adding fluoresceinated sheep anti-mouse immunoglobulin antibody for 30 min at 4 °C. Natural killer activity was assayed on ^{51}Cr-labelled YAC cells. Release of label into the supernatants was measured on a γ-counter and compared with background release and a positive control (nude mouse spleen cells).

* Cells were incubated overnight with latex particles and examined microscopically for evidence of phagocytosis.

† Sheep red blood cells (SRBCs) were washed in normal saline and resuspended to 5% (v/v) and mixed with an equal volume of 1:60 rabbit anti-SRBC antiserum. The cells were incubated for 1 h at 37°C, washed and resuspended to 2% v/v. Aliquots (50 μl) were added to 1 ml cultures of each cell line. Ig, immunoglobulin.

‡ Equal parts of 2% immunoglobulin-SRBCs (as in †) and cells at 5×10^6 per ml were incubated at 37°C for 15 min. The cells were centrifuged at 200g for 5 min and held at 4°C for a further 10 in. The suspension was stained with 0.1% crystal violet and rosettes counted.

§ The presence of membrane immunoglobulin and Thy-1 were tested for by staining with fluoresceinated rabbit anti-mouse immunoglobulin and fluoresceinated monoclonal anti-Thy-1.2 antibody (ref. 15) respectively.

‖ Secreted immunoglobulin was assayed using a sensitive radioimmunoassay in 96-well microtitre trays coated with rabbit anti-mouse immunoglobulin antiserum.

¶ A monoclonal antibody specific for a macrophage cell surface antigen[24] obtained from Dr E. Handman, The Walter and Eliza Hall Institute.

\# Esterase stains were performed using the α-naphthyl butyrate method (Sigma). Scoring: ±, occasional cells (<10%) showing faint black granules; ++, many cells (>75%) showing unequivocally positive granules. In the conditions used thymocytes and peripheral blood lymphocytes were negative or faintly positive, but peritoneal exudate cells showed strong positive granules.

** Macrophage activation by concanavalin A-stimulated spleen cell supernatants was assayed as previously described[16].

Table 2 Sodium butyrate induces increased differentiation of 5-AC-treated cells

	Control (%)	0.5 mM butyrate (%)	1 mM butyrate (%)
Morphological changes*	10 ± 3	96 ± 4	97 ± 5
Erythrophagocytosis†	4 ± 2	24 ± 6	7 ± 4
Esterase†	+	+ +	+ +

* The presence of granules and vacuoles similar to those seen in macrophages.
† Performed as described in Table 1 legend.

were removed at 0, 2 and 7 days and cultured at limit dilution in medium without 2-ME. No lines were obtained from the cells removed at day 0 or 2, but from day 7 cultures, 11 clones were obtained, 6 of which were morphologically similar to the 7 original lines. Both this and the original set of cell lines have been re-cloned and have remained stable for over 3 months of continuous culture. Final and compelling evidence that these cell lines were derived from ABLS 8.1 was that the pattern of rearrangement of immunoglobulin J_H genes in one of the derived cell lines was identical to the characteristic pattern seen in ABLS 8.1 (S. Cory, personal communication).

The granules and vacuoles present in the cell lines derived by 5-AC treatment of ABLS 8.1 conferred on the cells the superficial appearance of macrophages (Fig. 1). We performed several tests which indicated that, unlike ABLS 8.1, the derived cells possessed characteristics of macrophage cell lines (Table 1). The most impressive of these was phagocytosis of opsinized erythrocytes (Table 1, Fig 1) which was not seen in control ABLS 8.1 cells and seems to be a unique property of macrophages. Other strong supporting evidence was the presence of esterase-positive granules and the presence on a significant percentage of cells of an antigen detected by a macrophage-specific monoclonal antibody (Mac-1)[15]. As may be expected from the heterogeneity inherent even in cloned cell lines, these properties were not expressed by all cells in the population. Other properties such as the expression of Ia antigens and Fc receptors are less useful discriminants, as they are shared by both B lymphocytes and macrophages. Interleukin-1 (IL-1) production and the ability to present antigen are still being investigated. The cells did not produce detectable colony-stimulating factors (or T-cell growth factor or P-cell stimulating factor). They were not cytotoxic to a good target for natural killer cells (YAC cell line); nor did the derived cells lyse another tumour target (P815) in the presence of medium conditioned by concanavalin A-stimulated spleen cells known to contain macrophage activating factor[16].

In attempts to increase the proportion of more differentiated cells, agents known to induce differentiation were added to the cultures. Of those tested, dimethyl sulphoxide (0.5–10% v/v), actinomycin D (0.1–10 μg ml^{-1}), and tetradecanoyl phorbol acetate (1–100 ng ml^{-1}) or combinations of these had no effect at concentrations which did not severely affect viability. On the other hand, sodium butyrate (0.1–1 mM), while reducing the rate of cell division, did not obviously reduce viability in 3–5-day cultures. At concentrations of 0.3–1 mM, increasing proportions of differentiated cells were observed over 2–3 days (Table 2). After 3 days, in the presence of 0.5–1 mM sodium butyrate morphological changes such as the presence of granules and vacuoles had occurred in almost all the cells. In parallel experiments using ABLS 8.1, the presence of sodium butyrate was associated with the appearance of a small percentage (14%) of larger cells having an increased number of vacuoles. However, these cells were clearly distinguishable from those seen in 5-AC-treated cells cultured with butyrate, as they lacked granules and did not phagocytose erythrocytes.

We conclude that the demethylating drug 5-AC has induced heritable morphological and functional changes in the pre-B-cell line ABLS 8.1 characterized by the expression of macrophage-like features. The percentage of cells showing these properties was markedly increased by sodium butyrate which slowed the rate of cell division. This effect seemed to require prolonged exposure to the drug, which may suggest that a definite sequence or combination of demethylation steps is required to induce the observed changes. The finding that the only variants were macrophage-like cells may be explained in two ways. First, the culture conditions used might select for the survival of such cells. A second more interesting possiblity is that this result reflects the origin of both pre-B lymphocytes and macrophages from a common progenitor and that the demethylation activates certain genes which cause ABLS 8.1 to differentiate along a programmed or favoured path into a closely related, but functionally different cell. A close relationship between the lymphocyte and macrophage lineages is supported by evidence that the immunoglobulin μ-chain constant region gene is transcribed in myeloid lines and macrophages[17]. Although rearrangement of immunoglobulin variable (V) region genes has not been found in macrophage-like tumour lines examined so far, transcription of the μ gene in these cell lines may imply that the region of the genome encoding immunoglobulins undergoes some activation in a common precursor of B cells and macrophages. The fact that rearrangement of V genes has occurred on both chromosomes in ABLS 8.1 suggests that in this respect, ABLS 8.1 is at a B cell-specific differentiation stage beyond this putative common precursor. Nevertheless, given some activation of the immunoglobulin-coding region of the genome in macrophages[17,18] and, presumably, their precursors, it is possible that the reciprocal situation—some activation of macrophage-specific genes—also occurs in pre-B cells and accounts for the tendency for perturbation of the ABLS 8.1 genome by 5-AC to result in cells having macrophage characteristics. We are also investigating whether different cloning conditions could allow us to grow out other related haematopoietic cells.

Although experiments of this nature may provide useful information about the genetic mechanisms of gene expression, at the present stage of knowledge, extrapolation from the observed changes in phenotype to conclusions about physiological sequences of differentiation must be cautious. However, given the minimal amount known concerning the number and sequence of differentiation steps between multipotential haematopoietic stem cells and B cells[19], new approaches such as this may well yield useful clues to normal differentiation events.

We thank P. F. Bartlett, J. Adams, S. Cory, S. Gerondakis and A. W. Harris for helpful discussions and Ina Frangeskakis for technical assistance.

Received 21 December 1981; accepted 13 April 1982.

1. Groudine, M. & Weintraub, H. *Cell* **24**, 393–401 (1981).
2. Groudine, M., Eisenman, R. & Weintraub, H. *Nature* **292**, 311–317 (1981).
3. Naveh-Mang, T. & Cedar, H. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4246–4249 (1981).
4. Wu, C. *Nature* **284**, 856–860 (1980).
5. van der Ploeg, L. H. T. & Flavell, R. A. *Cell* **19**, 947–958 (1980).
6. McGhee, J. D. & Ginder, G. D. *Nature* **280**, 419–420 (1979).
7. Taylor, S. M. & Jones, P. A. *Cell* **17**, 771–779 (1979).
8. Pollack, Y., Stein, R., Razin, A. & Cedar, H. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6463–6467 (1980).
9. Wigler, M., Levy, D. & Perucho, M. *Cell* **24**, 33–38 (1981).
10. Jones, P. A. & Taylor, S. M. *Cell* **20**, 85–93 (1980).
11. Sklar, M. D., Shevach, E. M., Green, I. & Potter, M. *Nature* **253**, 550 (1975).
12. Premkumar, E., Potter, M., Singer, P. A. & Sklar, M. D. *Cell* **6**, 149 (1975).
13. Cory, S., Adams, J. M. & Kemp, D. J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4943–4947 (1980).
14. Wiener, F., Klein, G. & Harris, H. *J. Cell Sci.* **15**, 177–183 (1974).
15. Ledbetter, J. A. & Herzenberg, L. A. *Immun. Rev.* **47**, 63 (1979).
16. Pace, J. L. & Russell, S. W. *J. Immun.* **126**, 1863–1867 (1981).
17. Kemp, D. J., Harris, A. W., Cory, S. & Adams, J. M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2876–2880 (1980).
18. Kemp, D. J., Harris, A. W. & Adams, J. M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 7400–7404 (1980).
19. Schrader, J. W., Clark-Lewis, I. & Bartlett, P. F. in *Biology of Bone Marrow Transplantation*, 443–459 (Academic, New York, 1980).
20. Burgess A. W. *et al. Expl Haemat.* **9**, 893–903 (1981).
21. Schrader, J. W. & Clark-Lewis, I. *J. Immun.* **126**, 1101–1105 (1981).
22. Clark-Lewis, I. & Schrader, J. W. *J. Immun.* **127**, 1941–1947 (1981).
23. Ozato, K., Mayer, N. & Sachs, D. H. *J. Immun.* **124**, 533–540 (1980).
24. Springer, T., Galfré, G., Secher, D. S. & Milstein, C. *Eur. J. Immun.* **9**, 301–306 (1979).

HLA-DR light chain has a polymorphic N-terminal region and a conserved immunoglobulin-like C-terminal region

James F. Kaufman* & Jack L. Strominger

Department of Biochemistry and Molecular Biology, Sherman Fairchild Biochemistry Building, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA

The major histocompatibility complex (MHC) is a genetic region originally defined by graft rejection and now known to encode at least three classes of molecules which have important roles in the immune system. The MHC class I and class II antigens are both polymorphic two-chain cell-surface glycoproteins which are recognized by T lymphocytes. However, they are generally recognized by different subsets of T cells and have different functions, different tissue distributions and, by all available evidence, different structures[1-4]. Much is known about the detailed structure of the class I antigens (HLA-A, B, C in human; H-2K,D,L in mouse)[5-7]. The 44,000 (44K)-molecular weight (M_r) heavy chain consists of an amino-terminal extracellular region composed of three 10K M_r (90 amino acid) domains, a small hydrophobic membraneous segment and a small hydrophilic intracellular carboxy-terminal domain. The two amino-terminal domains are polymorphic, bear the carbohydrate and have no sequence homology with immunoglobulin. The third domain, closest to the membrane, and the 11.6K light chain (β_2-microglobulin) are highly conserved and have strong sequence homology with immunoglobulin. The structure of class II antigens (DR, DC1 in human; I-A, I-E in mouse) is much less well understood. Previous experiments involving papain proteolysis of native DR antigen have shown that both the light and heavy chains have a large glycosylated amino-terminal extracellular region, a small hydrophobic membranous region and a small carboxy-terminal hydrophilic region[8]. We have now applied limited proteolysis to demonstrate that the extracellular region of the light chain consists of two domains, each with a disulphide loop. The amino-terminal domain bears the carbohydrate and is polymorphic, while the carboxy-terminal domain is relatively conserved and has significant amino acid sequence homology with immunoglobulin. These results suggest a new picture of class II antigens indicating strong structural similarities to both class I antigens and immunoglobulins.

The B lymphoblastoid cell line WT51 (derived from an individual homozygous for HLA-DR4) was metabolically labelled with ^{35}S-cysteine. A detergent-solubilized cell lysate was prepared and treated with various proteases, the DR antigens immunoprecipitated with the rabbit xenoserum anti-p23, 30 and the precipitates analysed by SDS-polyacrylamide gel electrophoresis and fluorography. Before proteolysis, the DR4 antigen consists primarily of a 34K M_r heavy chain and a 29K M_r light chain (Fig 1*a*, lane 1). Proteolysis with chymotrypsin (Fig. 1*a*, lane 2) and with certain other proteases generates three major bands: 33K, 15.7K and a doublet of 12.6K and 13.2K. Proteolysis with trypsin or endoproteinase Lys-C (Fig. 1*a*, lanes 3, 4) generates a different pattern also with three major bands, of M_r 33K, 17K and 11.3K. Proteolysis with clostripain (Fig. 1*a*, lane 5) generates only two new bands at 33K and 27K.

Serial proteolysis with papain showed that the 33K and 27K bands represent the DR heavy and light chains with the carboxy-terminal hydrophilic regions removed. These fragments were generated by every protease tested, albeit with different efficiencies and at slightly different sites[9]. Only papain in very large amounts removes the hydrophobic regions[8], but the light chain could be further cleaved into large fragments by two classes of protease as shown in Fig. 1. Several experiments with DR4 antigen (including amino acid sequencing, see below) demonstrated that the larger bands (N_{ct}, 15.7K; N_t, 17K) are amino-terminal fragments of the light chain and the smaller bands (C_{ct}, 12.6K and 13.2K; C_t, 11.3K) are carboxy-terminal fragments of the light chain. Serial proteolysis with papain showed that the 13.2K band represents the entire carboxy-terminal chymotryptic fragment, while the 12.6K band represents the carboxy-terminal chymotryptic fragment lacking the hydrophilic carboxy-terminal peptide. Similar results were obtained with DRw6 antigen[9].

Endoglycosidase digestion has demonstrated that the DR heavy chain bears two N-linked glycans of $M_r \sim 3$K, whereas the light chain bears only one[10]. The antibiotic tunicamycin inhibits the glycosylation of DR antigens during biosynthesis (Fig. 1*b*, lanes 1, 3)[11]. Chymotryptic digestion of lysates from tunicamycin-treated cells generates two bands (26K and 13K) which immunoprecipitate with anti-p23,30, instead of the expected three bands (Fig. 1*b*, lanes 2, 4). Other experiments (not shown) using endoglycosidase D and antisera specific for either DR heavy or light chain demonstrate that the 26K fragment is the chymotryptic product of the non-glycosylated heavy chain, while the 13K band contains both the carboxy-terminal and the non-glycosylated amino-terminal fragments of the light chain. Thus, the amino-terminal chymotryptic fragment of the light chain bears the carbohydrate; also, the chymotryptic site is roughly in the middle of the chain. These results were confirmed with Drw6 antigen[9].

The mobility of denatured light chain in SDS-polyacrylamide gel electrophoresis depends markedly on whether or not it is reduced[12]. This mobility shift is greater than that of β_2-microglobulin (one disulphide loop), the heavy chain of HLA-A and -B antigens (two disulphide loops) and the heavy chain of DR antigens (one disulphide loop) (Fig. 2, lanes 1–4). Amino acid analyses had suggested that both the light and heavy chains of DR contained three cysteines[8,12]. However, the heavy chain could be labelled by mild reduction and alkylation, while the light chain could not[8]. Moreover, careful counting of cysteines in material from the lymphoblastoid line JY (DR4,w6) demonstrated that the heavy chain has one disulphide loop and one free sulphydryl, whereas the major light chain has four cysteines in two disulphide loops with no free sulphydryls[9]. It was thus of interest to determine the location of these disulphide loops in the light chain.

Lysates from ^{35}S-cysteine-labelled cells were proteolysed with either chymotrypsin or trypsin, immunoprecipitated with anti-p23,30, boiled in SDS sample buffer with or without 2 mM dithiothreitol (DTT) and analysed by SDS-gel electrophoresis. Both the amino-terminal and carboxy-terminal chymotryptic fragments contain comparable amounts of ^{35}S-cysteine and have a mobility shift equivalent to that of β_2-microglobulin (Fig. 2, lanes 5, 6). Thus, each chymotryptic fragment of the light chain contains a disulphide loop. This fact was confirmed by the trypsin cleavage. Electrophoresis in reducing conditions yielded the expected 17K amino-terminal and 11.3K carboxy-terminal tryptic fragments (Fig. 2, lane 7). Electrophoresis in non-reducing conditions yielded a band at approximately the mobility of the unreduced light chain (24K; Fig. 2, lane 8), corresponding to the amino-terminal and carboxy-terminal tryptic fragments still linked together by a disulphide bond. Thus, the trypsin cleaves between the two cysteines of the second disulphide loop, consistent with the fact that the amino-terminal fragment contains much more ^{35}S-cysteine than the carboxy-terminal tryptic fragment (Fig. 2, lane 7).

To define the cleavage fragments better, automated amino-terminal sequencing was carried out on chymotrypsin fragments from WT51 cells biosynthetically labelled with various amino acids (Fig. 3) The partial amino-terminal sequence for the 15.7K

* Present address: Basel Institute for Immunology, Grenzacherstrasse 487, Postfach CH-4005, Basel 5, Switzerland.

0028-0836/82/250694—04$01.00

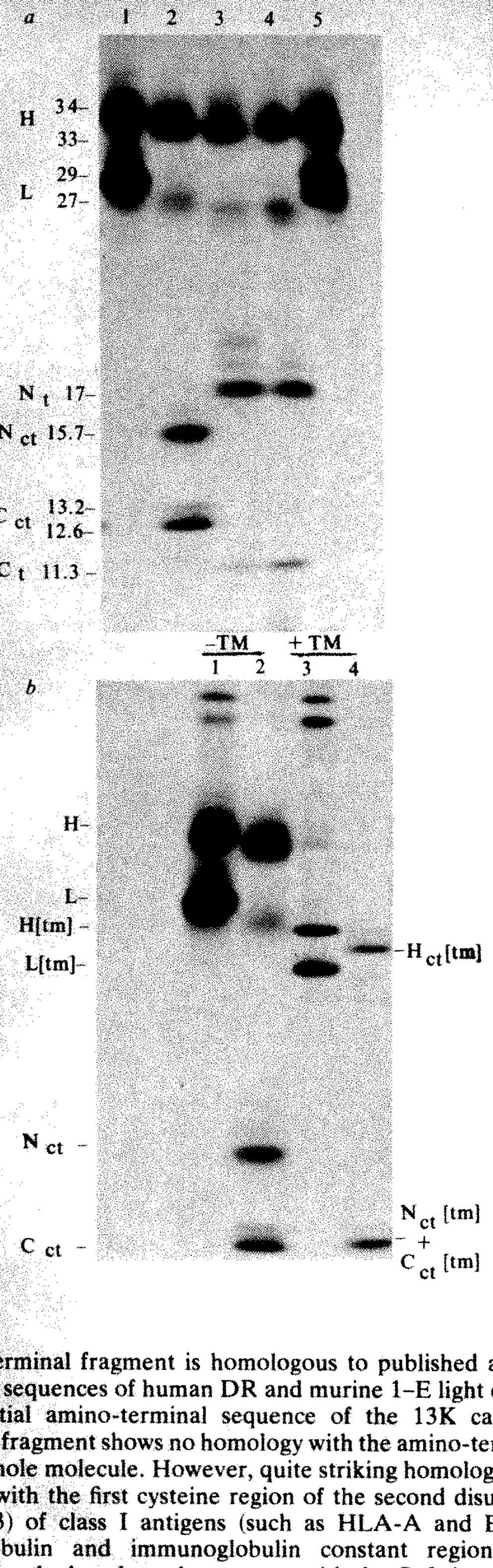


Fig. 1 SDS gel of DR antigen cleaved by various proteases. Except where noted, cells were labelled as follows. 2×10^6 WT51 cells were preincubated for 30 min at 37 °C in 2 ml cysteine-free RPMI-1640 medium supplemented with 10% dialysed fetal bovine serum (Gibco), 100 U ml^{-1} penicillin and 0.1 mg ml^{-1} streptomycin. Cells were centrifuged and resuspended in the same medium supplemented with 1 mCi ^{35}S-cysteine (NEN) and then incubated for 20 h at 37 °C. For tunicamycin-treated cells, 4 $\mu g\ ml^{-1}$ tunicamycin (TM) was included in both the preincubation and incubation media. Labelled cells were collected by centrifugation and incubated for 30 min on ice in 1 ml lysis buffer containing 2% Nonidet-P40 (NP40), 20 mM Tris-HCl *p*H 8, 1 mM $MgCl_2$, 150 mM NaCl and 0.1 mM phenylmethylsulphonyl fluoride (PMSF). Subcellular debris was removed by centrifugation in an Eppendorf centrifuge for 10 min at 4 °C and then the lysate was frozen at −80 °C until use. Except where noted, proteolysis was carried out as follows. One volume of lysate (usually 5 ul) was incubated with 10 volumes (50 µl) of buffer or protease at 37 °C for 1 h and then proteolysis was stopped by incubation on ice for 30 min with added proteolysis inhibitor. The serine proteases chymotrypsin (2 mg ml^{-1}, 40 U per mg, Worthington), diphenylcarbamylchloride-treated trypsin (2 mg ml^{-1}, 8,000 U per mg, Sigma), endoproteinase Lys-C (0.2 mg protease ml^{-1}, 0.3 U per mg, Boehringer-Mannheim) in 40 mM Tris-Cl *p*H 8, 100 mM NaCl, 10 mM $CaCl_2$ were halted by 0.2 volumes (1 µl) 100 mM PMSF in absolute ethanol and the sulphydryl proteases papain (0.01 mg ml^{-1}, 25 U per mg, Millipore) and clostripain (2 mg ml^{-1}, 40 U per mg, Boehringer-Mannheim) in 10 mM Tris-Cl *p*H 8, 0.1 mM EDTA, 1 mM DTT were halted by 5 volumes (25 µl) 25 mM iodoacetamide in water. Forty volumes (200 µl) of NET-NO (50 mM Tris-Cl *p*H 8, 5 mM EDTA, 150 mM NaCl, 0.5% NP40, 1 mg ml^{-1} ovalbumin) were added and the solution precleared with normal rabbit serum and formalin-fixed *Staphylococcus aureus* Cowan I strain (Staph A), immunoprecipitated with the rabbit xenoserum anti-p23,30[18] and Staph A, and washed as described previously[19]. Immunoprecipitates were eluted by boiling in sample buffer (200 mM Tris-Cl *p*H 8.8, 10% glycerin, 0.1% bromophenol blue, 5 mM EDTA, 2% SDS, 5% 2-mercaptoethanol) and analysed by electrophoresis through 7–15% polyacrylamide gradient gels (using Laemmli buffers[20] with 2 mM EDTA added) and fluorography[21]. In *a*, 5 µl lysate was incubated for 1 h at 37 °C with 50 µl buffer or protease and then proteolysis was halted on ice with either 1 µl 100 mM PMSF or 25 µl 25 mM iodoacetamide. 200 µl NET-NO was added and the sample was precleared, immunoprecipitated with anti-p23,30 and analysed. Lane 1, no protease; lane 2, chymotrypsin; lane 3, trypsin; lane 4, endoproteinase Lys-C; lane 5, clostripain. Numbers on the left represent M_r ($\times10^{-3}$). *b*: Lane 1, DR, no protease; lane 2, DR, chymotrypsin; lane 3, tunicamycin DR, no protease; lane 4, tunicamycin DR, chymotrypsin. H, DR heavy chain; L, DR light chain; N_t, amino-terminal tryptic fragment; N_{ct}, amino-terminal chymotryptic fragment; C_{ct}, carboxy-terminal chymotryptic fragment; C_t, carboxy-terminal tryptic fragment; tm, tunicamycin.

amino-terminal fragment is homologous to published amino-terminal sequences of human DR and murine I-E light chains. The partial amino-terminal sequence of the 13K carboxy-terminal fragment shows no homology with the amino-terminus of the whole molecule. However, quite striking homologies are evident with the first cysteine region of the second disulphide loop ($\alpha3$) of class I antigens (such as HLA-A and B), β_2-microglobulin and immunoglobulin constant regions. On inspection, the best homology occurs with the C_H3 domains of IgG (in descending order C_H3, C_H1, C_L, C_H2 and V).

The amino acids which are conserved or invariant within immunoglobulin domains are known to correspond to particular structural features[13,14]. The portion under consideration here includes the second and third antiparallel β-strands (found in the four-strand (fx) and three-strand (fy) sheets, respectively) and the loop between them (b2). In this region, the highly conserved residues (numbered from the site homologous to the chymotrypsin site in DR light chain) include the invariant cysteine-8 (involved in the disulphide loop), invariant tryptophan-22 (which shields the disulphide bond), highly conserved proline-15 (found in the bend) and the conserved alternating hydrophobic amino acids in the antiparallel β-strands (whose side chains fill the space between the two β-pleated sheets of the immunoglobulin domain: leucine-6, cysteine-8, valine-10, phenylalanine-13, isoleucine-18, valine-20, tryptophan-22). Every conserved or invariant residue in this region of the C_H3 domain of IgG myeloma Eu is identical to that of the carboxy-terminal chymotryptic fragment of the DR light chain. These same residues are conserved or identical in the second disulphide loop region of HLA-A and -B and in β_2-microglobulin, except that β_2-microglobulin has leucine-22 instead of the invariant tryptophan. Several other residues determined are conserved. In the aligned sequence, the chymotryptic site is located in the middle of the loop between the first and second antiparallel β-strands (b1). The trypsin site is likely to be located at lysine-30 (the first endoproteinase Lys-C-sensitive site after the chymotrypsin site), which is in the loop between the third and fourth anti-parallel β-strands (b3). These sites are at the same end of the immunoglobulin domain[13,14], and in the DR light chain, that end must be exposed to solvent. In fact, the chymotrysin site is accessible even in membrane-bound DR antigens (J.K., unpublished).

In class I antigens of both human and mouse, the regions with homology to immunoglobulin (β_2-microglobulin and the second disulphide loop of class I antigens) are invariant or highly conserved[5-7]. It was of interest to determine whether this is also true for the DR light chain, which is the polymorphic chain[15,16]. DR antigens from a number of cell lines (homozygous for DR1–8) were proteolysed with chymotrypsin, immunoprecipitated with anti-p23,30 and analysed by isoelectric focusing (IEF). By this criterion, the amino-terminal chymotryptic fragments of the light chain vary considerably (Fig. 4, lanes 1–8), whereas the carboxy-terminal fragments are quite conserved (Fig. 4, lanes 9–16). Whether the slight variability

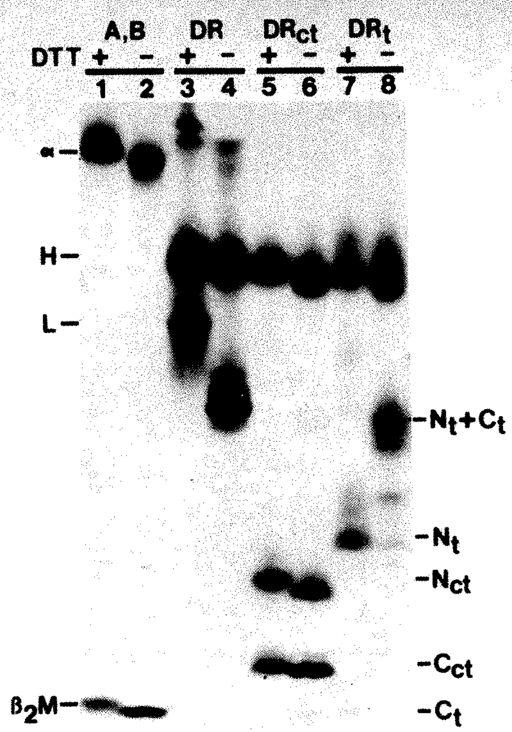


Fig. 2 SDS gel of reduced and nonreduced proteolytic fragments of DR antigen. As described for Fig. 1, samples were proteolysed and immunoprecipitated with anti-p23,30 or w6/32 (a monoclonal antibody which recognizes HLA-A, -B and -C antigens)[22]. The immunoprecipitates were boiled in sample buffer with either no reducing agent or 2 mM DTT and analysed by SDS-gel electrophoresis and fluorography. Lane 1, no protease, w6/32, DTT; lane 2, no protease, w6/32, no DTT; lane 3, no protease, anti-p23,30, DTT; lane 4, no protease, anti-p23,30, no DTT; lane 5, chymotrypsin, anti-p23,30, DTT; lane 6, chymotrypsin, anti-p23, 30, no DTT; lane 7, trypsin, anti-p23,30, DTT; lane 8, trypsin, anti-p23, 30, no DTT. α, HLA-A,B heavy chain; β_2M, β_2-microglobulin.

in carboxy-terminal fragments is due to polymorphism at the site of proteolysis or to limited polymorphism elsewhere (as in mouse β_2-microglobulin[17] or human $\alpha 3$[6]) is unclear. However, at least in the case of DRw6 from cell line LB and DR4 from cell line WT51, the chymotrypsin sites are homologous[9].

Protease cleavage of the native DR antigen has defined four domains for the DR light chain: a polymorphic glycosylated amino-terminal domain with a disulphide loop but no sequence homology with immunoglobulin, a relatively conserved domain with a disulphide loop and strong sequence homology with immunoglobulin, a hydrophobic region and a carboxy-terminal hydrophilic region. This organization of domains, as defined by protease cleavage, has been found in light chains from every class II antigen tested—human DR antigens (specificities 1–8), human DC1 antigens, mouse I-A and I-E antigens[9]. In fact, this general domain organization also extends to class II heavy chains. Papain proteolysis studies[8] and amino acid sequence data derived from genomic clones (A. Korman, personal communication) demonstrate that the DR heavy chain also has a four-domain structure: an amino-terminal domain having no structural homology with immunoglobulin, a second domain having striking sequence homology with immunoglobulin, a hydrophobic region and a carboxy-terminal hydrophilic region. The model based on these data (Fig. 5) allows homologous pairing between the domains of the two chains.

Like the class I antigens, the extracellular region of class II antigens consists of four domains: two having sequence homology to immunoglobulin and two without. It would be surprising if the gross three-dimensional structure and the function at the molecular level were different for the class I and II antigens. Another intriguing point is that the organization of domains in both the DR light and heavy chains is strikingly like that of a class I antigen heavy chain lacking one of the two

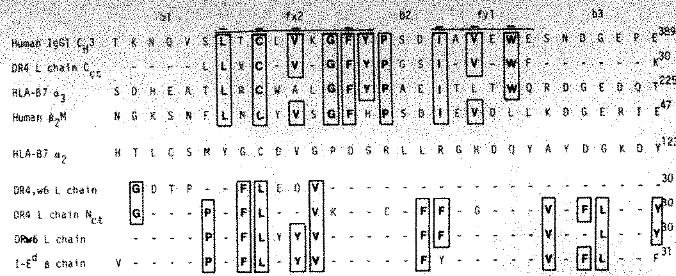


Fig. 3 Amino-terminal sequence of the carboxy-terminal chymotryptic fragment of the HLA-DR4 antigen. Each radioactive amino acid (the highest specific activity available from NEN) was lyophilized to near-dryness and then dissolved in RPMI-1640 medium lacking the appropriate amino acid. WT51 cells were labelled with individual radioactive amino acids -Ala, Cys, Phe, Ile, Lys, Leu, Met, Pro, Ser, Val, Trp, Tyr, 1–3 mCi ^{3}H, 100 μCi ^{14}C, 1–2 mCi ^{35}S per 2×10^6 cells) and a cell lysate prepared as outlined in Fig. 1 legend. (Cells were labelled with radioactive alanine in medium lacking both alanine and glutamic acid.) 0.5 ml lysate was incubated for 1.5 h at 37 °C with 5 ml of 2 mg ml^{-1} chymotrypsin and then proteolysis was halted with 50 μl 100 mM PMSF. 0.5 ml NET-NO was added and the lysate was cleared with 25 μl normal rabbit serum and 300 μl 10% Staph A. The precleared lysate was immunoprecipitated twice with 20 μl ascites of LB3.1 (a monoclonal antibody recognizing DR antigen) and 250 μl 10% Staph A. The immunoprecipitates were washed as usual and eluted by boiling into 90-μl sample buffer. A ^{14}C-labelled sample and or a ^{3}H-labelled sample were applied to the same 20 cm 7–15% polyacrylamide gradient SDS gel as a ^{35}S-cysteine-labelled sample. The presence of the ^{35}S-cysteine allowed the bands to be located by wet gel autoradiography after electrophoresis[16]. Gel slices were excised, crushed and eluted by shaking for 24 h at 4 °C into 1 ml of 0.1% SDS, 0.35 mg ml^{-1} sperm whale myoglobin (Beckman). The eluates were filtered through siliconized glass wool, lyophilized to 200 μl and acetone-precipitated. The precipitate was dissolved in 300 μl 50% formic acid and applied with 1.5 mg of Polybrene to the spinning cup of a Beckman 890C sequencer. The sample was subjected to automated Edman degradation using a 0.1 M Quadrol program[23], with double-coupling at the first step. Cycle fractions were transferred to counting vials, dried under nitrogen and the radioactivity determined using Aquasol and a Beckman liquid scintillation counter with standard ^{3}H and ^{14}C windows. Assignments of serine and glycine were made on the basis of samples labelled with ^{14}C-serine, in which the label is partly converted to glycine and should thus be considered tentative. The following residues from published sequences are shown: Eu (359–389)[24], human β-microglobulin (17–47)[25], HLA-B7 (93–123, 195–225)[26], I-E^d β-chain (1–31)[27] and DR light chains (1–30)[12,27]. Residues with identity to DR chymotryptic fragments are boxed. Overlined positions are located in the designated β-strand of Eu C_H3 and double overlined positions have amino acid side chains which fill the space between the β-pleated sheets. Residues are in the single letter code: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.

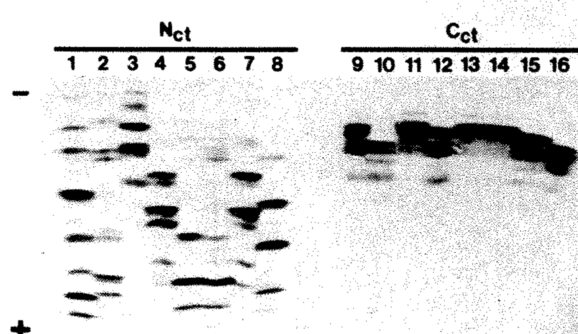


Fig. 4 IEF gel of chymotryptic fragments of DR antigens from various cell lines. Cells were labelled with ^{35}S-cysteine, proteolysed with chymotrypsin and immunoprecipitated with anti-p23,30 as described in Fig. 1 legend. After SDS-gel electrophoresis of samples, bands across all the lanes at the mobilities of the amino-terminal and carboxy-terminal fragments were excised (using dansylated β-lactoglobulin and myoglobin as markers). These gel slices were prepared for IEF by soaking in 8 M urea, 2% NP40, 0.2% pH 9–11 ampholines for 1 h at room temperature. The gel slices were applied to the acid end of a polymerized IEF gel containing 8 M urea, 2% NP40, 7% acrylamide, 0.19% N,N'-methylenebisacrylamide, 2% (w/v) ampholines (LKB) (pH 3.5–10, 4–6, 5–7, 6–8, 7–9, 9–11; 2:1:1:1:1:1) and focused at 1,000 V for 8 h at 7 °C using an LKB Multiphor 2117 flatbed gel apparatus[15]. Lanes 1–8, ^{35}S-cysteine-labelled DR, amino-terminal chymotryptic fragments; lanes 9–16, ^{35}S-cysteine-labelled DR, carboxy-terminal chymotryptic fragments. Lanes 1 and 9, MAJA (DR1,1); lanes 2 and 10, PGF (DR2,2); lanes 3 and 11, LKT (DR3,3); lanes 4 and 12, WT51 (DR4,4); lanes 5 and 13, MICH (DR5,5); lanes 6 and 14, LB (DRw6,w6); lanes 7 and 15, MANN (DR7,7); lanes 8 and 16, MADURA (DR8,8).

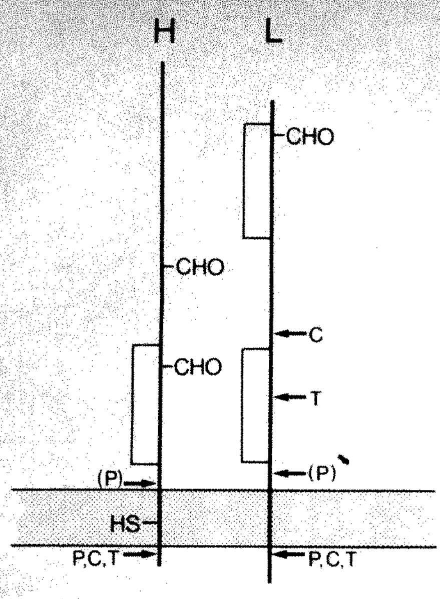


Fig. 5 Proposed model of HLA-DR antigen. H, heavy chain; L, light chain; P, papain site; C, chymotrypsin site; T, trypsin site. Thin lines represent disulphide bonds; -SH, free sulphydryl. Stippled area represents the membrane.

amino-terminal domains which do not have sequence homology with immunoglobulin. It therefore seems likely that both class I and II antigens descended from a common ancestral gene containing an immunoglobulin-like region.

We thank Dr R. Hamil (Lilly) for tunicamycin, Peter Knudsen for the gift of LB3.1 monoclonal antibody, Mike Krangel, Debbie Shackelford and Richard Pink for critical reading of the manuscript, members of the Strominger laboratory for endless discussions, and Cindy Lehn, Wendy Breisinger and Catherine Nordstrom for secretarial assistance. This research was supported by NIH grant AI10736.

Note added in proof: Since the submission of this paper, a preliminary report of the partial sequence of a human class II antigen has appeared[28]. The sequence is consistent with the data presented here.

Received 5 January; accepted 23 April 1982.

1. Klein, J. *The Biology of the Mouse Histocompatibility 2 Complex* (Springer, New York, 1975).
2. Bodmer, W. F., Batchelor, J. R., Bodmer, J. G., Festenstein, H. & Morris, P. J. (eds) *Histocompatibility Testing* 1977 (Munksgaard, Copenhagen, 1978).
3. Klein, J. *Science* **203**, 516–521 (1979).
4. Dorf, M. E. (ed.) *The Role of the Major Histocompatibility Complex in Immunobiology* (Garland, New York, 1981).
5. Strominger, J. L. *et al.* in *The Role of the Major Histocompatibility Complex in Immunobiology* (ed. Dorf, M. E.) 115–172 (Garland, New York, 1981).
6. Ploegh, H. T., Orr, H. T. & Strominger, J. L. *Cell* **24**, 287–299 (1981).
7. Coligan, J. E., Kindt, T. J., Uehara, H., Matrinko, J. & Nathenson, S. G. *Nature* **291**, 35–39 (1981).
8. Kaufman, J. F. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **76**, 6304–6308 (1979).
9. Kaufman, J. F. & Strominger, J. L. *J. Immun.* (submitted).
10. Shackelford, D. A. & Strominger, J. L. *J. biol. Chem.* (submitted).
11. Korman, A. J., Ploegh, H. L., Kaufman, J. F., Owen, M. J. & Strominger, J. L. *J. exp. Med.* **152**, 65s–82s (1980).
12. Springer, T. A., Kaufman, J. F., Terhorst, C. & Strominger, J. L. *Nature* **268**, 213–218 (1977).
13. Beale, D. & Feinstein, A. *Q. Rev. Biophys.* **9**, 135–180 (1976).
14. Anzel, L. M. & Poljak, R. J. *A. Rev. Biochem.* **48**, 961–997 (1979).
15. Shackelford, D. A. & Strominger, J. L. *J. exp. Med.* **151**, 144–165 (1980).
16. Kaufman, J. F., Andersen, R. L. & Strominger, J. L. *J. exp. Med.* **152**, 37s–53s (1980).
17. Gates, F. T., Coligan, J. E. & Kindt, T. J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 554–558 (1981).
18. Humphreys, R. E. *et al. J. exp. Med.* **144**, 98–112 (1976).
19. Kessler, S. W. *J. Immun.* **115**, 1617–1624 (1975).
20. Laemmli, U. K. *Nature* **227**, 680–685 (1970).
21. Bonner, W. M. & Laskey, R. A. *Eur. J. Biochem.* **46**, 83–88 (1974).
22. Barnstable, C. J. *et al. Cell* **14**, 9–20 (1978).
23. Brauer, A. W., Margolis, M. N. & Haber, E. *Biochemistry* **14**, 3029–3035 (1975).
24. Edelman, G. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **63**, 78–85 (1969).
25. Peterson, P. A., Cunningham, B. A., Berggard, I. & Edelman, G. M. *Proc. natn. Acad. Sci. U.S.A.* **69**, 1697–1701 (1972).
26. Orr, H. T., Lopez de Castro, J. A., Lancet, D. & Strominger, J. L. *Biochemistry* **18**, 5711–5720 (1979).
27. Allison, J. P. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 3953–3956 (1978).
28. Kratzin *et al. Hoppe-Seylers Z. physiol. Chem.* **362**, 1665–1669 (1981).

Expression of lymphocyte surface IgE does not require switch recombination

Yoshio Yaoita, Yoshihiro Kumagai*, Ko Okumura* & Tasuku Honjo

Department of Genetics, Osaka University Medical School, Osaka 530, Japan
* Department of Immunology, Faculty of Medicine, University of Tokyo, Tokyo 113, Japan

Immunoglobin heavy (H) chains are composed of a variable (V_H) and a constant (C_H) region. The latter is encoded respectively by eight distinct genes for the classes and subclasses in mice—C_μ, C_δ, $C_{\gamma3}$, $C_{\gamma1}$, $C_{\gamma2b}$, $C_{\gamma2a}$, C_ϵ and C_α genes—arrayed in that order on the chromosome[1]. During differentiation of a single B lymphocyte, a given V_H region is first expressed as a μ chain, followed by the switch of the C_H region to other classes such as δ, γ, ϵ and α. The molecular genetic basis for this phenomenon, called heavy chain class switch, has been elucidated recently by cloning and characterization of immunoglobulin genes of mouse myelomas secreting various classes of immunoglobulin[2–4]. By this model, DNA rearrangement, called S–S recombination, brings a V_H gene, located originally 5′ to the C_μ gene, close to another C_H gene by deletion of an intervening DNA segment[5–9]. The S–S recombination occurs between S regions located in the 5′ flanking region of each C_H gene. The nucleotide sequences of S regions comprise tandem repetitive sequences sharing short common sequences[10–14]. In contrast, the expression of surface IgD in $\mu^+\delta^+$ lymphomas and in normal $\mu^+\delta^+$ lymphocytes seems not to involve DNA rearrangement in the region between the C_μ and C_δ genes[15,16]. The simultaneous expression of the C_μ and C_δ genes with a single V_H gene may be mediated by two alternative routes of RNA processing of a primary nuclear transcript containing the V_H, C_μ and C_δ genes. We have now studied the organization of C_H genes in sorted $\mu^+\epsilon^+$ B lymphocytes and found that they retain C_μ, C_δ, C_γ and C_ϵ genes, suggesting that the simultaneous expression of the C_μ and C_ϵ genes is mediated by an RNA splicing mechanism. We propose that class switching requires at least two steps of differentiation, the first step involving activation of differential splicing and the second the DNA rearrangement.

During experiments to define the new mouse C_H locus encoding IgE (*Igh-7*)[17], we found that the *Igh*-congeneic strain SJA/9 (*Igh^a*), which has a SJL background, cannot produce a detectable amount of IgE in the sera even after infection with *Nippostrongylus brasiliensis*, which is known to stimulate polyclonal IgE production. Furthermore, we found that the increase of IgE-bearing B cells after *N. brasiliensis* infection occurs equally in SJA/9 and SJL mice (K.O. *et al.*, in preparation). Approximately 10% of the spleen cells of the *N. brasiliensis*-infected mice carry IgE on their surface, providing a unique opportunity for identifying sufficient IgE-bearing B lymphocytes for molecular genetic analyses. The advantage of SJA/9 mice is that a low IgE level in sera minimizes the binding of IgE to Fc receptors of ϵ-negative lymphocytes, thus avoiding the contamination of ϵ-negative cells into sorted ϵ^+ B cells. This strategy—taking advantage of a low level of a certain isotype in sera—was successfully used to purify IgG2a-bearing B cells from allotype-suppressed mice by fluorescence activated cell sorter[18].

The IgE-bearing B cells were isolated from spleen cells of *N. brasiliensis*-infected SJA/9 mice using the fluorescence-activated cell sorter. Only the brightest top 9% of the stained cells were collected and their purity examined under the fluorescent microscope, which is less sensitive and gives a lower limiting value of staining. As shown in Table 1, at least 86% of the sorted cells were brightly stained with anti-ϵ antibody. Most of the ϵ-bearing cells also carried the μ chain on their surface.

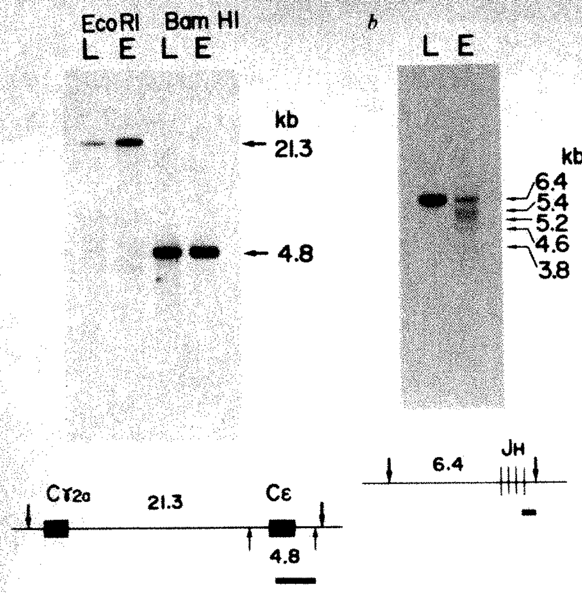


Fig. 1 Analysis of *Eco*RI and *Bam*HI fragments of ε-bearing cell and SJA/9 liver DNAs using cloned mouse C_ε and J_H genes as probes. DNA was isolated from sorted ε^+ B cells according to the method of S. Gattoni (personal communication). About 8×10^6 cells were lysed in 0.5 ml of 0.1% SDS, 10 mM EDTA 10 mM Tris-HCl *p*H 8.0, 10 mM NaCl, 200 μg ml^{-1} proteinase K for 90 min at 30°C with gentle shaking. An equal amount of neutralized heated (60 °C) phenol was added to the lysate and tilted for 15 min. A few millilitres of ethanol were added to the water phase and DNA was wound up at the interphase with a glass rod. DNA was washed in a large amount of ethanol and dissolved in 100 μl H_2O. About 6 μg DNA was obtained. Liver DNA was isolated as described before. The fragments obtained by *Eco*RI and *Bam*HI digestions of ε-bearing cell DNA (E) and SJA/9 liver DNA (L) were fractionated by electrophoresis in a 0.5% agarose gel, transferred to nitrocellulose filters and hybridized to the cloned mouse C_ε (*a*) and J_H (*b*) probes which are indicated by wide bars below the restriction maps. In *b*, only *Eco*RI digests were analysed. Each lane contains ~1 μg DNA. Probes were labelled by nick-translation and hybridized as described[9]. In the restriction maps closed rectangles indicate structural genes. Numbers show lengths (kb) of fragments. ↓, *Eco*RI; ↑, *Bam*HI.

However, they were not stained with anti-δ, anti-γ2a or anti-γ1 antibody (data not shown). As a control, about 8% of the ε-depleted cells and 10% of the whole spleen cells were stained with anti-ε antibody. The results indicate that the sorted cells are the essentially pure population of $\mu^+\varepsilon^+$ B cells.

To confirm that IgE on the surface of $\mu^+\varepsilon^+$ B cells is endogenously synthesized, ε^+ B cells were treated with trypsin (2.5 mg ml^{-1}) for 30 min at 37°C to strip off all cell-surface immunoglobulins, then after culturing, newly synthesized immunoglobulins on the surface were re-examined by fluorescence staining. As expected, 2 and 5 h after the trypsin treatment, 84 and 97%, respectively, of cells were stained with anti-ε. Furthermore, the sorted ε^+ B cells of SJA/9 were shown to secrete IgE when T cells of SJL were provided (K.O. *et al.*, in preparation).

We have extracted DNA from the sorted $\mu^+\varepsilon^+$ cells and examined the C_H gene organization in ε-bearing cells using the Southern blotting technique. When DNA of ε^+ B cells was digested with *Eco*RI, blotted and hybridized with the C_ε gene probe, it produced a 21.3-kilobase (kb) fragment identical to that produced in SJA/9 liver DNA (Fig. 1*a*). Similarly, *Bam*HI digestion of DNAs of ε^+ B cells and SJA/9 liver yielded an identical fragment (4.8 kb) hybridizing with the C_ε probe. As the *Eco*RI fragment (21.3 kb) encompasses the whole region between the $C_{\gamma2a}$ and C_ε genes, the above results indicate that the C_ε gene does not rearrange in the IgE-bearing lymphocytes, unlike the IgE-secreting hybridoma and myeloma[19,20].

On the other hand, the J_H gene fragment of the IgE-bearing lymphocyte DNA drastically reduced the intensity as compared with that of SJA/9 liver DNA, and appeared blurred, in agreement with the interpretation that a large number of different rearrangements have generated many new *Eco*RI fragments of different lengths in polyclonal B cells[21] (Fig. 1*b*). Each of several faint bands (5.4, 5.2, 4.6 and 3.8 kb) having less than a few per cent of the intensity of the germ-line band in liver DNA, may represent rearranged J_H genes for anitbodies against *N. brasiliensis* antigens *per se*.

Table 1 Characterization of sorted ε-bearing cells

Cells	Surface immunoglobulin-positive cells (%) μ	ε
Whole spleen cells	46	9.6
Sorted cells		
Sorted ε^+ B cells	77	86
The remaining cells	45	8.3

SJA/9 mice were infected with *N. brasiliensis* by subcutaneous injection of third-stage larvae (750 per mouse). The mice were killed 2 weeks later and their spleens gently teased to obtain lymphocytes. Spleen lymphocytes were stained with guinea pig anti-murine IgE sera[17] and fluorescein isothiocyanate-labelled rabbit anti-guinea pig IgG1 antibodies. Stained spleen cells were sorted on the FACS III (Becton-Dickinson). An aliquot of the sorted cells and spleen cells were stained with biotin-conjugated rabbit anti-murine μ antibodies and rhodamine isothiocyanate-labelled avidin after photobleaching. As stained cells were counted under the fluorescent microscope, the values shown are lower limits.

When DNA of IgE-bearing cells was digested with *Eco*RI, blotted and hybridized with the $C_{\gamma2b}$–$C_{\gamma2a}$ probe, it produced the 6.6- and 21.3-kb fragments which correspond to the germ-line forms of the $C_{\gamma2b}$ and $C_{\gamma2a}$ genes, respectively (Fig. 2*a*). The germ-line $C_{\gamma2b}$ and $C_{\gamma2a}$ gene fragments (9 and 6.2 kb, respectively) were also detected in DNA of IgE-bearing cells when digested with *Hin*dIII. Similarly, *Eco*RI or *Hin*dIII digestion of IgE-bearing cell DNA produced the germ-line forms of the $C_{\gamma1}$ and $C_{\gamma3}$ genes (data not shown). In fact, the faint 23-kb *Hin*dIII fragment is the germ-line $C_{\gamma1}$ gene detected by cross-hybridization of the $C_{\gamma2b}$ probe (Fig. 2*a*).

*Eco*RI digestion of the ε^+ B-cell DNA produced the germ-line form of the C_μ gene fragment (13 kb) (Fig. 2*b*). Inasmuch as the 13-kb C_μ fragment contains the whole S_μ region, there

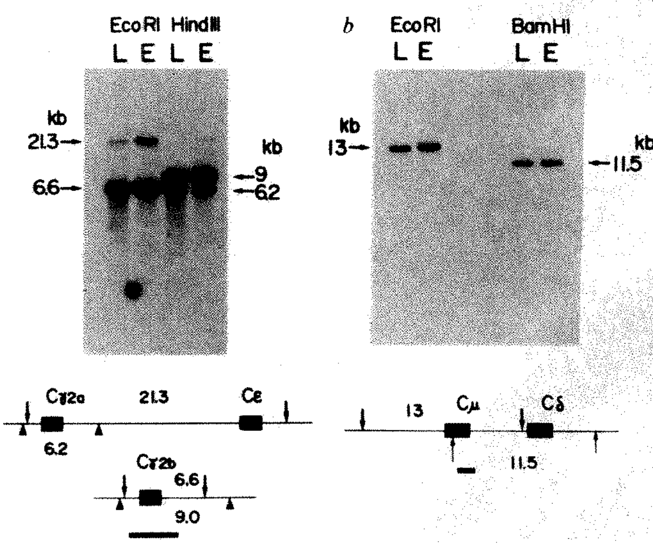


Fig. 2 Analysis of *Eco*RI, *Hin*dIII and *Bam*HI fragments of ε-bearing cell and SJA/9 liver DNAs using cloned mouse $C_{\gamma2b}$ and C_μ genes as probes. Experimental conditions are as described in Fig. 1 legend. The fragments produced by *Eco*RI, *Hin*dIII and *Bam*HI digestion of ε-bearing cell DNA (E) and SJA/9 liver DNA (L) were electrophoresed and blotted to nitrocellulose filters. The restriction maps surrounding probes used [$C_{\gamma2b}$ (*a*) and C_μ (*b*)] are shown below. Each lane contains ~1 μg DNA. ▲, *Hin*dIII; ↓, *Eco*RI; ↑, *Bam*HI.

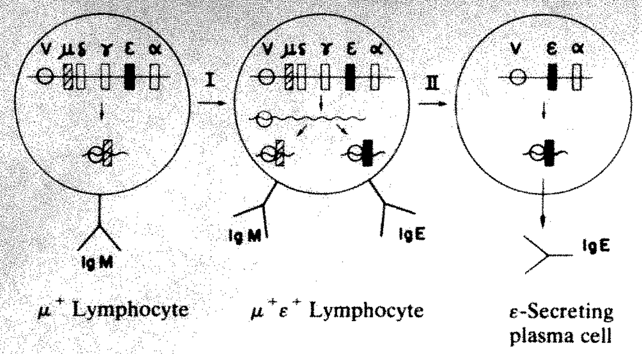


Fig. 3 Two steps of differentiation from μ^+ lymphocytes to ε-secreting plasma cells. Step I involves the activation of differential splicing. Alternate splicing of a long transcript containing V_H, C_μ, C_γ and C_ε sequences will produce mRNA encoding either μ or ε chain with the same V-region sequence. Step II involves DNA deletion. See the text.

is no doubt about the absence of the DNA rearrangement in the S_μ region. *Bam*HI digestion of the ε^+ B-cell and SJA/9 liver DNAs yielded the 11.5-kb fragment hybridizing with the C_μ probe. As the 11.5-kb *Bam*HI fragment encompasses the C_δ as well as the C_μ gene, the results indicate that the C_δ gene is not rearranged in the ε^+ B cells.

The intensity of each C_H gene band was quantified at the corresponding size by hybridizing the same filters used in the above experiments with the β-globin and ribosomal RNA gene probes. The ratio of the intensity of each C_H gene band to the β-globin and ribosomal RNA gene bands as determined by densitometer tracings remained constant regardless of the origin of DNA, that is, from liver or ε^+ B cells. These results suggest that the organization of the C_H gene in the IgE-bearing cells is probably the same as the germ-line gene except that the J_H is rearranged. These data indicate that IgE expression in $\mu^+\varepsilon^+$ B cells does not involve the S_μ–S_ε rearrangement although it is well established that the intervening DNA segment is deleted by the S–S recombination in various immunoglobulin-secreting B-cell lymphomas, hybridomas and myelomas including IgE secreters[5–9].

Given these results we propose that differentiation of IgM-bearing B lymphocytes to IgE-secreting plasma cells may proceed by at least two biochemical steps, as shown in Fig. 3. The first step (step I) promotes differentiation of IgM-bearing B lymphocytes into IgM–IgE-bearing B lymphocytes, which involves the activation of differential RNA processing of a single large RNA transcript containing V_H, C_μ, C_δ, C_γ and C_ε gene sequences. The large transcript may be spliced into μ or ε mRNA by specific enzymes and/or specific assisting molecules such as low molecular weight RNA[22,23]. The size of the primary transcript is estimated to be ~180 kb from the C_H gene organization[1]. Naturally, the μ and ε mRNAs share an identical V_H region sequence. As we handled a mixed populaton of $\mu^+\varepsilon^+$ lymphocytes, we were unable to determine whether the same V_H sequence was associated with the C_μ and C_ε sequences in a single cell. However, IgD and IgM molecules were shown to bear the identical V_H region in a $\mu^+\delta^+$ lymphoma[16]. We presume that step I does not involve any major DNA rearrangement.

IgM–IgE-bearing B lymphocytes differentiate into IgE-secreting B cells or plasma cells by step II, which involves S_μ–S_ε recombination and simultaneous DNA deletion as established previously[2–9]. Obviously, a similar mechanism should apply to the switch from IgM-bearing cells to IgD-, IgG- or IgA-secreting plasma cells. It is not clear whether DNA rearrangement accompanies the differentiation from IgM-bearing B cells to IgM-secreting plasma cells. Interestingly, most of the IgM-secreting myelomas and hybridomas seem to have a deletion in the S_μ region[7,24,25]. We think it reasonable that deletion of the S_μ region facilitates the transcription of the C_μ gene and promotes IgM secretion as the extremely G+C-rich S_μ region[11] may hinder efficient transcription.

There are no data concerning the length of the primary transcript of μ mRNA in IgM-bearing B cells. They may transcribe the whole C_H gene locus from the beginning. If so, step I is mediated by the activation of a new differential splicing system. Alternatively, the primary transcript in IgM-bearing B cells may contain only the V_H and C_μ sequences. In this case, step I requires at least two new biochemical events—the transcription of a much larger RNA and the activation of a new differential splicing system. To avoid the premature termination of transcription, lymphocytes may have to introduce some biochemical changes in the C_H gene locus such as demethylation[26]. In fact, the C_δ gene is demethylated in $\mu^+\delta^+$ hybridoma but not in μ^+ lymphoma[27].

This model favours the hypothesis that the splicing as well as recombination mechanism is class specific. Otherwise, the isotype expression in B cells should be transient and multiple (more than three isotypes per cell) until they become plasma cells. Several lines of evidence suggest that the expression of a certain V_H sequence is closely associated with a specific C_H isotype. CBA/N mice have genetic defects which make them incapable of producing anti-phosphorylcholine antibody of any classes other than IgE whereas anti-phosphorylcholine antibody of IgM and IgG is very common in most mouse strains[28,29]. A lymphoma cell line I.29 has been shown consistently to switch from μ to α[30]. Such results seem to indicate that the S–S recombination is catalysed by the class-specific enzyme(s).

IgM–IgE-bearing lymphocytes accumulated in spleens of *N. brasiliensis*-infected SJA/9 mice are capable of differentiating into IgE-secreting plasma cells when T cells of SJL are provided (K.O. *et al.*, in preparation). As SJA/9 mice can synthesize normal amounts of IgM, IgG and IgA, the defect of SJA/9 seems to reside in IgE-specific regulatory T cells[31]. Furthermore, it is probably at step II that the T cells affect B-cell differentiation.

After completion of this manuscript we learned that Perlmutter and Gilbert[32] had found the C_μ gene in γ_1-bearing B cells purified from normal spleen using antibody-coated Petri dishes.

We thank Drs T. Tada and T. Kishimoto for critical reading of the manuscript and encouragement, Y. Sakagami and S. Nishida for technical assistance, and F. Oguni for preparing this manuscript. This investigation was supported in part by grants from the Mitsubishi Science Foundation, the Torey Science Foundation, the Naito Foundation and the Ministry of Education, Science and Culture of Japan.

Received 23 March; accepted 21 April 1982.

1. Shimizu, A. *et al. Cell* (in the press).
2. Kataoka, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **77,** 919–923 (1980).
3. Davis, M. M. *et al. Nature* **283,** 733–739 (1980).
4. Sakano, H. *et al. Nature* **286,** 676–683 (1980).
5. Honjo, T. & Kataoka, T. *Proc. natn. Acad. Sci. U.S.A.* **75,** 2140–2144 (1978).
6. Cory, S. *et al. Nature* **285,** 450–456 (1980).
7. Coleclough, C. *et al. Proc. natn. Acad. Sci. U.S.A.* **77,** 1422–1426 (1980).
8. Rabbitts, T. *et al. Nature* **283,** 351–356 (1980).
9. Yaoita, Y. & Honjo, T. *Nature* **286,** 850–853 (1980).
10. Kataoka, T. *et al. Cell* **23,** 357–368 (1981).
11. Nikaido, T. *et al. Nature* **292,** 845–848 (1981).
12. Nikaido, T. *et al. J. biol. Chem.* (in the press).
13. Davis, M. M. *et al. Science* **209,** 1360–1365 (1980).
14. Dunnick, W. *et al. Nature* **286,** 669–675 (1980).
15. Moore, K. W. *et al. Proc. natn. Acad. Sci. U.S.A.* **78,** 1800–1804 (1981).
16. Maki, R. *et al. Cell* **24,** 353–365 (1981).
17. Borges, M. S. *et al. Immunogenetics* **13,** 499–507 (1981).
18. Okumura, K. *et al. Eur. J. Immun.* **6,** 467–472 (1976).
19. Nishida, Y. *et al. Proc. natn. Acad. Sci. U.S.A.* **78,** 1581–1585 (1981).
20. Nishida, Y. *et al. Proc. natn. Acad. Sci. U.S.A.* (in the press).
21. Nottenberg, C. & Weissman, I. L. *Proc. natn. Acad. Sci. U.S.A.* **78,** 484–488 (1981).
22. Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. & Steitz, J. A. *Nature* **283,** 220–224 (1980).
23. Rogers, J. & Wall, R. *Proc. natn. Acad. Sci. U.S.A.* **77,** 1877–1879 (1980).
24. Hurwitz, J. L. *et al. Cell* **22,** 349–359 (1980).
25. Yaoita, Y. & Honjo, T. *Biomed. Res.* **1,** 164–175 (1980).
26. Razin, A. & Riggs, A. D. *Science* **210,** 604–610 (1980).
27. Rogers, J. & Wall, R. *Proc. natn. Acad. Sci. U.S.A.* **78,** 7497–7501 (1981).
28. Sher, I. *et al. J. exp. Med.* **141,** 788–803 (1975).
29. Kishimoto, T. *et al. J. Immun.* **123,** 1039–1043 (1979).
30. Sitia, R., Rubartelli, A. & Hammerling, U. *J. Immun.* **127,** 1388–1394 (1981).
31. Kishimoto, T. in *Progress in Allergy* (eds Kallos, P., Ishizaka, K. & Waksman, B.) (Karger, Basel, in the press).
32. Perlmutter, A. P. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* (in the press).

Binding of four repressors to double-stranded *tet* operator region stabilizes it against thermal denaturation

Wolfgang Hillen & Bernhard Unger

Institut für Organische Chemie und Biochemie, Technische Hochschule Darmstadt, Petersenstrasse 22, D-6100 Darmstadt, FRG

The investigation of protein–DNA interactions benefits from methods for the dissection of chromosomal DNA into handy fragments and their subsequent preparation in large amounts[1,2]. Of particular interest are proteins active in gene regulation and their interaction with the control sequences of the respective genes[3]. Recently, we reported[4] the purification of the molecular components of the control elements from the tetracycline-resistance (*tet*) gene located on the transposon Tn*10*. The Tet repressor inhibits transcription of the *tet* gene and its own gene[4]. When the *tet* operator was prepared on a 187 base pair (bp) DNA fragment[4], the Tet repressor was found to bind specifically to this fragment with a stoichiometry of four Tet repressors per DNA fragment[4]. Tetracycline inhibits this binding[4] and operates *in vivo* as an inducer for the expression of the Tn*10*-encoded tetracycline resistance[5,6]. We now report thermal denaturation experiments of the Tet repressor–*tet* operator complex and demonstrate independently that four Tet repressor molecules bind to the 187 bp DNA and stabilize a 125 bp double-stranded DNA sequence against thermal denaturation.

Figure 1 summarizes the location of the biologically active sequences on the 187 bp DNA fragment. We investigated the thermal stability of the Tet repressor–*tet* operator complex by melting the 187 bp DNA in the presence of various amounts of Tet repressor. A stabilization can only be observed when the ionic conditions in the experiment are such that the thermal denaturation of the free DNA fragment occurs at a lower temperature than the denaturation of the complex. Previous results have suggested that the Tet repressor–*tet* operator complex is stable up to about 65 °C (ref. 4) at 0.2 M NaCl, 0.02 M $MgCl_2$. Assuming that the stability of the complex is limited by the thermal stability of the Tet repressor, then the melting experiments must be conducted at less than 40 mM NaCl because the T_m of the 187 bp fragment is 65 °C at that salt concentration. If, however, the report that non-

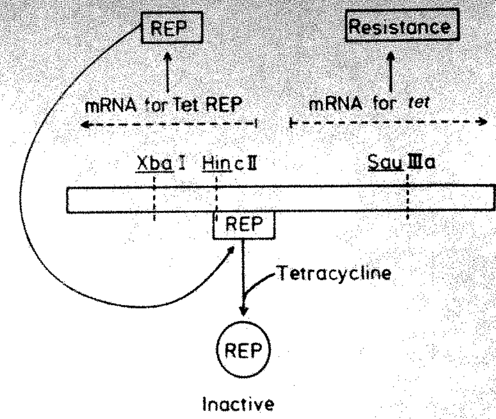


Fig. 1 Schematic presentation of the biologically active loci on the 187 bp *tet* operator DNA fragment (represented by the box). The recognition sites for the restriction endonuclease *Xba*I, *Hinc*II and *Sau*IIIa are indicated[6]. The *Hinc*II site is protected from cleavage by the presence of the Tet repressor (REP), whereas the *Xba*I and the *Sau*IIIa sites are not[4]. These results indicate the location of the *tet* operator region. The dashed lines represent the mRNAs for the tetracycline-resistance protein and the Tet repressor. The Tet repressor in the *tet* operator binding conformation is drawn as a box. It is inactivated in the presence of tetracycline as indicated by the circle[4].

specific binding of the lac repressor to DNA increases with decreasing ionic strength[7,8] were also to hold for the interaction of the Tet repressor with DNA, it would be difficult to distinguish nonspecific from specific binding. Figure 2 compares the effect of the Tet repressor on the thermal denaturation of two DNA fragments. Figure 2*a* shows the melting curves of the 95 bp DNA fragment containing the *Escherichia coli* lactose genetic control elements with (dashed line) and without (solid line) Tet repressor[9,10]. This experiment serves as a control for nonspecific binding. The data reveal that the T_m of the 95 bp fragment is increased by 3.2 °C in the presence of the Tet repressor. This effect may be caused by either nonspecific binding of the Tet repressor to the DNA, or a slight increase in the ionic strength resulting from addition of the protein solution, or a combination of the two. As the ionic strength is only 6 mM in this experiment, even small alterations could cause this elevated T_m.

Figure 2*b* shows the dependence of thermal denaturation of the 187 bp *tet* operator fragment on Tet repressor concentration. In the absence of Tet repressor the 187 bp DNA melts in a single cooperative transition (solid line), whereas in the presence of a twofold molar excess of Tet repressor over 187 bp DNA, three melting transitions can be distinguished (dashed

Table 1 Melting temperatures and area analysis of the thermal denaturation experiments

Concentration of Tet repressor ($M \times 10^7$)	Transition I T_m(°C)	Transition I % Area	Transition II T_m(°C)	Transition II % Area	Transition III T_m(°C)	Transition III % Area	Ratio of area III : area II
187 bp *tet* operator DNA fragment							
0	56.8	100	—	—	—	—	—
1.1	58.3	70	59.5	12	65.3	18	1.5
2.4	59.3	36	60.5	27	66.3	37	1.4
3.6	60.8	18	61.8	33	67.5	49	1.5
4.7	—	—	62.0	39	67.5	61	1.6
9.3	—	—	61.8	39	67.3	61	1.6
95 bp *lac* operator DNA fragment							
0	62.4	100	—	—	—	—	—
6.3	65.6	100	—	—	—	—	—

The total concentration of the 187 bp DNA fragment was kept constant at $\sim 1.1 \times 10^{-7}$ M and the 95 bp DNA fragment at 1.6×10^{-7} M. Other experimental conditions were as indicated in Fig. 2 legend. The accuracy of T_m determination is approximately ±0.3 °C as determined from repeat experiments. The total area was normalized to 100%. The accuracy of the area determination is approximately ±8% of the given value. This error arises mainly from the lack of distinction between peaks I and II (Fig. 2*b*). The number of base pairs in transition III was determined from melting experiments at 272 nm and is 125±7 bp.

0028-0836/82/250700–03$01.00

line). A van't Hoff analysis[11,12] reveals that transition I arises from denaturation of the remaining free 187 bp fragment. Transitions II and III are the result of the thermal denaturation of the Tet repressor–187 bp DNA complex. The dissociation rate constant of the repressor–operator complex must be low, because it takes 100 min to go from transition I to transition III in this experiment (see Fig. 2 legend). In these conditions, the operator sequence is stabilized by ~7 °C in the complex. In the presence of a fourfold molar excess of Tet repressor over 187 bp DNA (dotted line), all the 187 bp fragment melts as the protein–DNA complex in transitions II and III.

Table 1 summarizes the results from this experiment, carried out with a range of Tet repressor concentrations, together with the data from the 95 bp melting experiments. Three conclusions may be drawn. (1) Whereas the area of subtransition I decreases and those of peaks II and III increase with increasing Tet repressor concentration, the ratio of the areas under peaks II and III remains constant, indicating that peak II always results from the denaturation of 62±7 bp and peak III from that of 125±7 bp. For the conversion of peak areas to number of base pairs, the total area of the melting transitions of a given species is set to 100%. Each subtransition then arises from the denaturation of the respective percentage of the total number of base pairs[12]. It may therefore be concluded that the same species of protein–DNA complex is denaturing at all Tet repressor concentrations studied, which excludes contributions from nonspecific binding. The shift of denatured base pairs from peak I to peaks II and III with increasing Tet repressor concentration is the result of the higher proportion of the Tet repressor–operator complex. (2) With increasing amounts of Tet repressor, the T_m values of the subtransitions are shifted to higher temperature. The increase in T_m of transition I is probably the result of increasing ionic strength due to the added protein solution. As, at low Tet repressor concentrations, all Tet repressors bind specifically, nonspecific binding cannot account for the increased T_m. Transition III represents the stabilization resulting from specific binding of the Tet repressor to the *tet* operator region. This specific stabilization is strong enough to disrupt the cooperativity of the melting process. Unoccupied sequences melt under transition II. They are more stable than the free DNA melting in transition I because they are next to the stabilized repressor–operator complex. Therefore, the single strands cannot dissociate, and this accounts for the difference in T_m between transitions I and II. (3) The T_m of peak III does not increase over 67.5 °C. This temperature must therefore be the denaturation temperature of the Tet repressor–*tet* operator complex. Thermal denaturation experiments with increasing ionic strength (data not shown) reveal that the stability of the complex is hardly affected by the salt concentration.

Based on this interpretation, the analysis of the melting curves in the presence of increasing amounts of Tet repressor should yield a binding curve for the Tet repressor–*tet* operator complex. If the fraction of DNA bound to the Tet repressor is calculated from an area analysis of peaks II and III with the areas of peak I representing the fraction of free DNA, the binding curve displayed in Fig. 3 is derived. The result indicates a stoichiometry of 3.6 Tet repressor molecules per DNA fragment, which, considering the uncertainties involved in determining the active fraction in a given protein preparation, may be interpreted as a 4:1 stoichiometry in the Tet repressor–*tet* operator complex.

In the *E. coli* lactose genetic control region, the stability of the DNA is correlated to the genetic functions of the region[9,10]. The binding sites for cyclic AMP receptor protein and RNA polymerase are located in one cooperatively melting 80 bp region of the DNA, on the boundary of which is the lac repressor

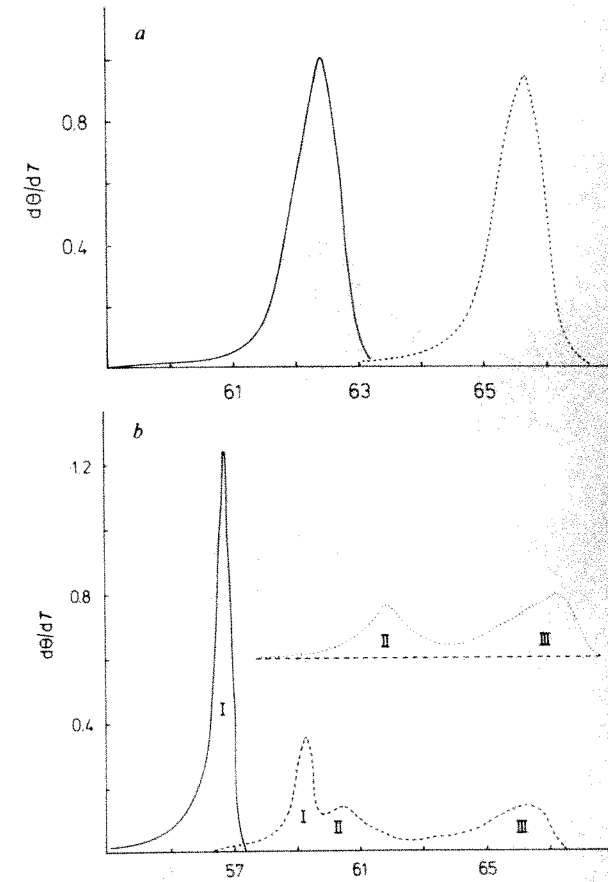


Fig. 2 Thermal denaturation of the 95 bp *lac* operator and the 187 bp *tet* operator DNA fragments in the presence of various amounts of Tet repressor. *a*, Melting transition of the 95 bp *lac* operator DNA fragment alone (—) and in the presence of a fourfold excess of Tet repressor (-----). The only effect of the Tet repressor is a shift of the T_m from 62.4 to 65.6 °C. In both experiments, the concentrations of the 95 bp fragment and the Tet repressor were 1.6×10^{-7} M and 6.3×10^{-7} M respectively. The shift of T_m is interpreted as a nonspecific effect arising from addition of the protein solution. *b*, Melting transitions of the 187 bp *tet* operator DNA fragment in the presence of various amounts of Tet repressor. The 187 bp DNA alone (—) melts in a single cooperative transition designated I. The concentration of the 187 bp fragment in all experiments was 1.1×10^{-7} M. In the presence of 2.4×10^{-7} M Tet repressor, three melting transitions are detectable (-----). Transition I results from the remaining free 187 bp fragment. Nonspecific effects probably arising from salts present in the Tet repressor solution shift the T_m by 2.5 °C. Results from the experiment in the presence of 4.7×10^{-7} M Tet repressor (·····) suggest that transitions II and III arise from the denaturation of the Tet repressor–*tet* operator complex. This melting curve is offset by 0.6 on the vertical axis. In these conditions, no free 187 bp DNA is left and the entire denaturation occurs under peaks II and III. The amount of active Tet repressor in the protein preparation used for this experiment was determined by titration with ^{3}H tetracycline assuming a stoichiometry of 1:1 for the Tet repressor–tetracycline complex as described previously[4]. The salt conditions in the thermal denaturation experiments were 3.5 mM NaCl, 5 mM sodium cacodylate, *p*H 7.0, and 0.1 mM EDTA. Electrophoresis of the samples on 5% acrylamide gels before and after the thermal denaturation experiments confirmed the absence of any degradative activity during the measurements. The thermal denaturation was measured using a Zeiss PMQ III spectrophotometer at 260 nm. The cuvette was heated by a Haake thermal bath equipped with a Haake PG 10 temperature controller. The heating rate was 0.07 °C per min. The temperature was monitored by placing a NTC resistor in the circulating liquid behind the cuvette. The Zeiss and the NTC were interfaced to a Commodore computer model 3016. Between 180 and 250 temperature–absorbance data pairs were recorded per experiment. The data were smoothed by fitting segments of 13–21 data pairs to a linear equation. Less rigorous smoothing procedures ensured the absence of any effect of the smoothing procedure on the melting curve. The total hypochromicity was normalized to 1 and the melting curve was differentiated to produce the figures.

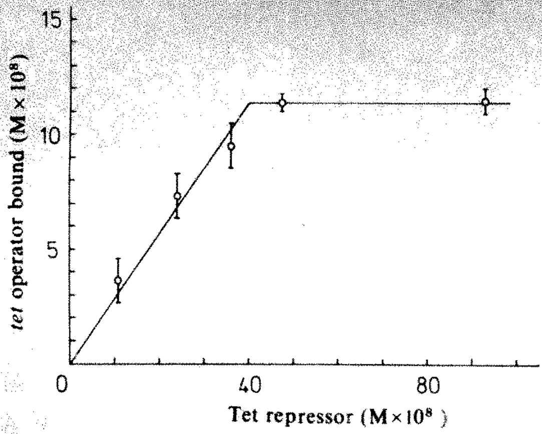


Fig. 3 Binding curve of the 187 bp *tet* operator DNA fragment to the Tet repressor. The areas of peak I (Table 1) for different Tet repressor concentrations were analysed to yield the concentration of bound 187 bp DNA, which was plotted versus the total Tet repressor concentration. The resulting binding curve indicates quantitative binding in these conditions. The stoichiometry obtained from this curve is ~3.6 Tet repressors per 187 bp fragment.

binding site[9,10]. A similar arrangement has been reported for the *tet* gene regulatory region on Tn*10*, where the operators and promoters are located on a 140 bp cooperatively melting part of the DNA[13]. Our demonstration here of the specific stabilization of this operator–promoter sequence by the Tet repressor suggests that transcription is regulated by a change in the stability of the double strand. In this view, a repressor would be expected to increase the stability of the double strand, which we have indeed shown to be the case, whereas an activator would have the opposite effect.

The data presented here suggest that the Tet repressor recognizes a double-stranded conformation of the *tet* operator which is considerably stabilized in the complex. A possible cruciform structure of the *tet* operator would involve single-stranded regions of the DNA, but such regions could not serve as a Tet repressor binding site because a lower T_m would be predicted for a single-stranded binding protein complexed to DNA[14]. The thermal denaturation of the Tet repressor–*tet* operator complex serves as a new method to monitor specific complex formation because the free and bound DNA is clearly distinct and the dissociation rate constant is low enough for the fraction of each to be determined[7]. The constant stoichiometry during the titration implies strong cooperativity of complex formation. Cooperativity in repressor–operator binding has also been observed for the λ repressor[15]. Because there may be two operator regions on the 187 bp fragment[6], it remains to be shown whether this cooperativity is the result of protein–protein contacts or of a conformational change in the DNA on binding the first Tet repressor.

We thank Dr H. G. Gassen for stimulating discussions and continued support and Mrs E. Rönnfeldt for helping prepare the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft.

Received 12 October 1981; accepted 13 April 1982.

1. Hardies, S. C. *et al. J. biol. Chem.* **254**, 5527–5534 (1979).
2. Hillen, W., Klein, R. D. & Wells, R. D. *Biochemistry* **20**, 3748–3756 (1981).
3. Wells, R. D. *et al. Prog. Nucleic Acid Res. molec. Biol.* **24**, 167–267 (1980).
4. Hillen, W., Klock, G., Kaffenberger, I., Wray, L. V. & Reznikoff, W. S. *J. biol. Chem.* (in the press).
5. Yang, H., Zubay, G. & Levy, S. *Proc. natn. Acad. Sci. U.S.A.* **73**, 1509–1512 (1976).
6. Wray, L. V., Jorgensen, R. A. & Reznikoff, W. S. *J. Bact.* (in the press).
7. Wang, A. C., Revzin, A., Butler, A. P. & van Hippel, P. H. *Nucleic Acids Res.* **4**, 1579–1593 (1977).
8. Record, M. T., Jr., de Haseth, P. L. & Lohmann, T. M. *Biochemistry* **16**, 4791–4802 (1977).
9. Hillen, W., Goodman, T. C., Benight, A. S., Wartell, R. M. & Wells, R. D. *J. biol. Chem.* **256**, 2761–2766 (1981).
10. Hardies, S. C., Hillen, W., Goodman, T. C. & Wells, R. D. *J. biol. Chem.* **254**, 10128–10134 (1979).
11. Riesner, D. & Römer, R. *Physico-Chemical Properties of Nucleic Acids* Vol. 2 (ed. Duchesne, J.) 237–318 (Academic, New York, 1973).
12. Hillen, W., Goodman, T. C. & Wells, R. D. *Nucleic Acids Res.* **9**, 415–436 (1981).
13. Hillen, W. & Unger, B. *Nucleic Acids Res.* (in the press).
14. Alberts, B. M. & Frey, L. *Nature* **227**, 1313–1315 (1970).
15. Johnson, A. D. *et al. Nature* **294**, 217–223 (1981).

Spherical *E. coli* due to elevated levels of D-alanine carboxypeptidase

Zdzisław Markiewicz*‡, Jennifer K. Broome-Smith†, Uli Schwarz* & Brian G. Spratt†

* Max-Planck-Institut für Virusforschung, Abteilung Biochemie II, D-7400, Tübingen, FRG
† Microbial Genetics Group, School of Biological Sciences, University of Sussex, Falmer, Brighton BN1 9QG, UK

The characteristic shape of bacterial cells is maintained by the strength and form of the murein layer of the cell wall[1]. Thus, rod-shaped bacteria need to synthesize a rod-shaped murein. The shape of the murein may be determined by the properties of the penicillin-binding proteins (PBPs) that catalyse the insertion of murein precursors into the cell wall[2]. In *Escherichia coli* inhibition of PBPs produces characteristic effects on bacterial morphology. For example, inactivation of PBP 2 results in an inability to synthesize cylindrical murein during cell elongation and the growth of *E. coli* as spherical cells[2]. We report here that osmotically stable spherical cells of *E. coli* can also be produced by an increase in the level of PBP 5, a D-alanine carboxypeptidase[3], and show that these cells, and those produced by inactivation of PBP 2, have the same abnormality in the structure of the newly inserted murein.

PBP 5 of *E. coli* is one of the major D-alanine carboxypeptidases that remove the terminal D-alanine from the pentapeptide side chains of the murein[2,3]. The PBP 5 gene (*dacA*) is carried by λ*dlip5*[4]. DNA from λ*dlip5* was digested partially with *Sau*3A and fragments of 1–5 kilobases (kb) were purified and inserted into the *Bam*HI site in the tetracycline resistance gene of the low copy number plasmid pSC105[4]. Plasmids carrying *dacA* were detected by the reappearance of PBP 5 in an *E. coli* strain deleted of the chromosomal *dacA* gene[6]. pBS25, which has a 3-kb chromosomal insert, was chosen for further study, and pBS32, which is pSC105 with a deletion of the tetracycline resistance gene, was used as a control plasmid. Figure 1 shows the elevated level of PBP 5 in C600 (pBS25) compared with C600 (pBS32) control cells.

E. coli C600 (pBS25) grew as spherical cells in all media tested (Fig. 2) and plated with low efficiency on nutrient agar but plated normally on minimal media. To establish that the spherical growth of *E. coli* was a direct result of the elevated level of PBP 5, rather than of the product of some other gene on pBS25, we subcloned the *dacA* gene on a 1.5-kb *Eco*RI–*Bam*HI fragment into a derivative[7] of pSC105. *E. coli* C600 carrying the resulting plasmid (pBS42) also grew as spherical cells. Insertion of DNA fragments into unique *Bst*EII, *Sal*I and *Bgl*II sites within the 1.5-kb *dacA* fragment inactivated *dacA* and in each case these plasmids had no effect on cell morphology. Similarly, a series of deletions extending from the *Eco*RI or the *Bam*HI ends of the *dacA* fragment in pBS42 were constructed, and in all cases plasmids with deletions extending into the *dacA* gene had no effect on cell morphology, whereas deletions that left *dacA* intact produced a spherical morphology. Spherical cells therefore result from the elevated levels of the *dacA* gene product, PBP 5.

E. coli cells that are rod-shaped at 30 °C and become spherical at 42 °C have been obtained by placing the expression of the *dacA* gene under the control of the phage λ leftward promoter and a thermolabile λ repressor, in a low copy number plasmid expression vector[7]. This strain eventually dies at 42 °C, presumably because the high levels of PBP 5 remove all the

‡ Present address: Institute of Microbiology, Warsaw University, Nowy Swiat 67, Poland.

Table 1 Murein metabolism in spherical cells produced by elevated levels of PBP 5

Strain	Cross-linkage: New murein	Cross-linkage: Old murein	Pentapeptide in new murein (%)	D-alanine carboxypeptidase activity (U)
C600(pBS32)	25.3±1.3	32.4±1.9	2.5	10.1
C600(pBS25)	30.2±1.7	32.8±1.7	1.5	37.8

Cross-linking in old murein: cells were grown at 37 °C for four generations in minimal medium containing kanamycin (50 μg ml^{-1}) and ^{3}H-diaminopimelic acid (2.5 μCi ml^{-1}; 50 Ci mmol^{-1}) to an absorbance at 578 nm of 0.6. Cross-linkage in new murein: cells were grown at 37 °C in minimal medium containing kanamycin (50 μg ml^{-1}) to an absorbance at 578 nm of 0.6 and samples (4 ml) were labelled with ^{3}H-diaminopimelic acid (30 μCi ml^{-1}) for 4 min. The labelled cells were processed as described previously[8], except that digestion with lysozyme was preceded by a 6-h incubation of the murein with amylase (100 μg ml^{-1}), and the level of the cross-linkage was determined as described previously[8]. Values given are the average of eight separate experiments. The pentapeptide content of the newly inserted urein was determined as described elsewhere[8]. In each of four independent experiments the pentapeptide level in C600(pBS25) was 48–73% of that in C600(pBS32), although the variation between experiments was rather large. Crude extracts from sonicated cells suspended in 0.5 M Tris-HCl buffer (pH 8.6) containing 2% Triton X-100 were used to determine carboxypeptidase activity which was assayed by measuring the release of ^{14}C-D-alanine from UDP-N-acetylmuramyl pentapeptide labelled in the terminal D-alanine (specific activity 21 c.p.m. per pmol). The reaction mixture contained 0.5 nmol labelled substrate, 0.1 M Tris-HCl buffer (pH 8.6), 0.4% Triton X-100 and extract in a volume of 25 μl. The reaction was terminated by boiling and product and substrate were separated by ascending paper chromatography (Whatman 3MM paper) in a solvent system of isobutyric acid: 1 M ammonia (5:3; vol/vol). One unit of carboxypeptidase is defined as that amount of enzyme that will release 1 pmol of ^{14}C-D-alanine from the substrate in 30 min at 37 °C.

terminal D-alanine residues from murein precursors and prevent further murein synthesis. Probably for the same reason *dacA* cannot be cloned into high copy number plasmid vectors.

The level of PBP 5 in C600 (pBS25) was about 10 times that in control cells. The level of total D-alanine carboxypeptidase in C600 (pBS25), which is due to PBPs 4, 5 and 6 (ref. 2), was 3.7 times higher than in control cells (Table 1). The elevated level of PBP 5 resulted in a decreased amount of pentapeptide side chains in the newly inserted murein (Table 1), supporting the view that PBP 5 acts as a D-alanine carboxypeptidase *in vivo*[8].

Cross-linking of the murein of *E. coli* occurs in two stages[9]. Initial insertion of murein precursors by transpeptidation, catalysed by high molecular weight PBPs[2,10,11], results in a low level of cross-linking in the newly inserted murein. Subsequently, further cross-linking of the inserted murein occurs to produce the level of cross-linking characteristic of the mature wall. As expected, the level of cross-linking in the newly inserted murein of control cells of C600 (pBS32) was lower than in old murein (Table 1). The level of cross-linking in the old murein of the spherical cells of C600 (pBS25) was the same as in the control cells, but the level in newly inserted murein was abnormally high, and was almost the same as in the old murein (Table 1).

The abnormal level of cross-linkage in the newly synthesized murein of the spherical cells of C600 (pBS25) was also found in *E. coli* K-12 cells made spherical by inactivation of PBP 2. In three separate experiments the newly synthesized murein was ~20% more cross-linked after growth in the presence of

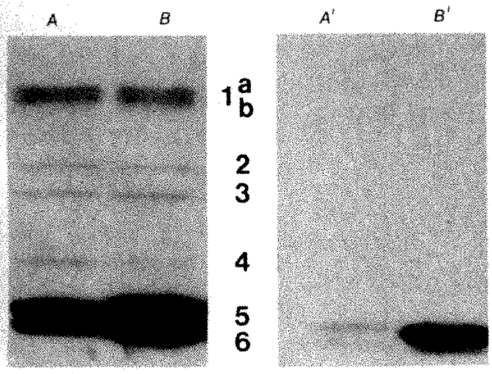


Fig. 1 Overproduction of PBP 5 in *E. coli* C600 (pBS25). C600 (pBS32) (*A* and *A'*) and C600 (pBS25) (*B* and *B'*) were grown to late exponential phase in minimal medium containing kanamycin (50 μg ml^{-1}) at 37 °C and cell envelopes were prepared and PBPs assayed as described previously[4]. The left and right halves of the figure show the same gel autoradiographed for 30 days and 3 days, respectively.

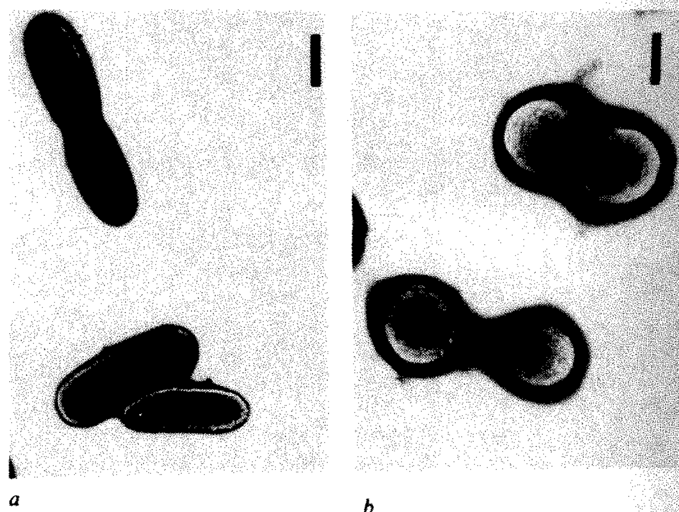


Fig. 2 Morphology of *E. coli* C600 (pBS25). Electron micrographs of negatively stained C600 (pBS32) (*a*); and C600 (pBS25) (*b*). Cells were grown at 37 °C in minimal medium containing kanamycin (50 μg ml^{-1}). Scale bars, 1 μm.

mecillinam (a specific inhibitor of PBP 2; ref. 12) than in control cells. The effect on cross-linking occurred within a few minutes of the addition of mecillinam and well before any shape change could be seen (Fig. 3). Similar results were obtained using the temperature-sensitive mutant, SP137, which becomes spherical at 42 °C due to inactivation of a thermolabile PBP 2 (ref. 12). The cross-linking in new murein was 30% higher in SP137 cells shifted to 42 °C for 40 min than in rod-shaped cells of SP137 maintained at 30 °C.

These experiments demonstrate that the production of spherical cells of *E. coli* by two completely different methods results in the same abnormality of murein synthesis. How does inactivation of PBP 2, or overproduction of PBP 5, result in a higher level of cross-linking in newly inserted murein and growth as spherical cells? It is unlikely that PBP 2 acts directly by modulating expression of PBP 5, or vice versa, as no changes in PBP 2 activity have been found in cells with elevated PBP 5 and no changes in PBP 5 activity accompany inactivation of PBP 2.

One possibility is that PBP 2 is a component of the enzyme system that incorporates new murein for cell elongation, and that in its absence murein synthesis occurs entirely via the system responsible for the synthesis of hemispherical murein during cell division. The latter system involves PBP 3, which

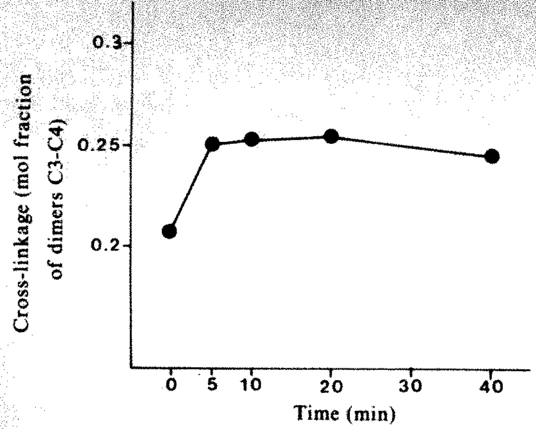


Fig. 3 Effect of mecillinam on cross-linking of newly inserted murein. *E. coli* PA3092[8] was grown in minimal medium at 37 °C to an absorbance at 578 nm of 0.4 and was pulse-labelled with ^{3}H-diaminopimelic acid (25 μCi ml^{-1}) for 4 min, before addition of mecillinam, or at various times after the addition of mecillinam (1 μg ml^{-1}). The cross-linking in newly inserted murein was measured as described in Table 1 legend.

seems to be a murein synthetase–transpeptidase[13], and may insert murein precursors into the wall with a higher degree of cross-linking than the cell elongation system. Overproduction of PBP 5 could result in spherical cells because the excessive removal of terminal D-alanine residues from murein precursors may convert them to a form which cannot be utilized by the murein system involved in cell elongation, but which is the preferred substrate for the cell division system. For example, the cell division system might require precursors with tetrapeptide side chains that act as acceptors in the cross-linking reaction, as occurs in *Gaffkya homari*[14], whereas, in cell elongation, precursors with pentapeptide side chains may be required to act as peptide donors in cross-linking. Increased conversion of pentapeptide side chains to tetrapeptide side chains in murein precursors would stimulate cross-linking by the cell division system, and suppress cross-linking by the cell elongation system, resulting in spherical growth.

Interestingly, the suggestion that D-alanine carboxypeptidase levels may be implicated in the switching between cell elongation and cell division has been proposed from completely different experiments that showed a high level of carboxypeptidase immediately before cell division and a low level in cells that elongate without dividing[15,16].

This work was supported by MRC project grant G978/99/C (to B. G. S.). We thank D. Mirelman for a gift of labelled UDP-*N*-acetylmuramyl pentapeptide, and J. Deeley, H. Frank and J. Sauter for technical assistance and electron microscopy.

Received 19 March; accepted 29 April 1982.

1. Rogers, H. J. *Adv. microb. Physiol.* **19,** 1–62 (1979).
2. Spratt, B. G. *Phil. Trans. R. Soc.* **B289,** 273–283 (1980).
3. Amanuma, H. & Strominger, J. L. *J. biol. Chem.* **255,** 11173–11180 (1980).
4. Spratt, B. G, Boyd, A. & Stoker, N. G. *J. Bact.* **143,** 569–581 (1980).
5. Cohen, S. N., Chang, A. C. Y., Boyer, H. W. & Helling, R. B. *Proc. natn. Acad. Sci. U.S.A.* **70,** 3240–3244 (1973).
6. Spratt, B. G. *J. Bact.* **144,** 1190–1192 (1980).
7. Stoker, N. G., Fairweather, N. F. & Spratt, B. G. *Gene* (in the press).
8. De Pedro, M. A., Schwarz, U., Nishimura, Y. & Hirota, Y. *FEMS Microbiol. Lett.* **9,** 219–221 (1980).
9. De Pedro, M. A. & Schwarz, U. *Proc. natn. Acad. Sci. U.S.A.* **78,** 5856–5860 (1981).
10. Matsuhashi, M. *et al.* in *Beta-Lactam Antibiotics* (ed. Mitsuhashi, S.) 203–223 (Springer, Berlin, 1981).
11. Suzuki, H. *et al. FEBS Lett.* **110,** 245–249 (1980).
12. Spratt, B. G. *J. Antimicrob. Chemother.* **3,** Suppl. B, 13–14 (1977).
13. Ishino, F. & Matsuhashi, M. *Biochem. biophys. Res. Commun.* **101,** 905–911 (1981).
14. Hammes, W. P. *Eur. J. Biochem.* **70,** 97–106 (1976).
15. Mirelman, D., Yashouv-Gan, Y., Nuchamovitz, Y., Rozenhak, S. & Ron, E. Z. *J. Bact.* **134,** 458–461 (1978).
16. Mirelman, D., Yashouv-Gan, Y. & Schwarz, U. *J. Bact.* **129,** 1593–1600 (1977).

In vitro packaging of plasmid DNAs into ΦX174 bacteriophage capsid

Akira Aoyama & Masaki Hayashi

Department of Biology, University of California at San Diego, La Jolla, California 92093, USA

Previously we have developed an *in vitro* system that synthesizes infectious single-stranded (ss) DNA bacteriophage ΦX174 from purified phage components[1]. This system provides an assay to examine the function of each viral component in the phage synthesizing reaction. In this report, we show that the specificity of the template DNA for the phage synthesizing reaction is determined by the specific region of ΦX174 genome. Plasmid DNAs carrying the origin of ss DNA synthesis of ΦX174 are active as templates in the *in vitro* system and produce particles containing ss plasmid DNA packaged into ΦX174 capsids. The orientation of the ΦX174 DNA fragment carrying the origin determines which strand of the plasmid is packaged. The particles synthesized *in vitro* are infectious to ΦX174-sensitive *Escherichia coli* and introduce ss plasmid DNA into the host cells. The ss DNA is converted to ds plasmid which is maintained in the cell.

In *E. coli* cells infected with ΦX174, ss DNA is first converted to double-stranded (ds) replicative form (RF) DNA, which is then multiplied by semiconservative replication[2]. During the late stage of infection, asymmetric ss DNA synthesis occurs[2,3]. This DNA synthesis is coupled to the encapsidation of ss DNA into preformed proheads to form infectious phage particles. Both semiconservative replication of RF DNA and asymmetric ss DNA synthesis are initiated by the nicking action of ΦX174 gene A protein at a specific site (A site) on the viral strand of ΦX174 RFI DNA[4,5]. The A site also acts as a signal for terminating these DNA syntheses[6]. However, it is not known whether the remainder of the ΦX174 sequence has some role in ss DNA synthesis and packaging of the ss DNA into prohead. This problem has been approached by cloning the A-site sequence from ΦX174 RF DNA into plasmid pACYC 184 (ref. 7) thereby isolating it from other ΦX174 DNA sequence and testing its template activity in the *in vitro* phage synthesizing system.

The third largest fragment of restriction endonuclease *Hin*cII-digested ΦX174 RF DNA (4,200 to 4,812 of the Sanger *et al.*'s map)[8], which contains the A-site sequence, was inserted at the *Eco*RI site of plasmid pACYC184 using *Eco*RI linkers. Plasmid DNAs containing the fragment in either direction were isolated (pH24R and pH24L). Plasmid pH13, which contains fragment III from a restriction endonuclease *Hae*III digest of ΦX174 RF DNA (4,948 to 434 of the Sanger *et al.*'s map) inserted at the *Eco*RI site of pACYC184, was also isolated for the control experiment.

The *in vitro* phage synthesizing system described previously[1] was used to examine the template activity of these plasmid DNAs (Table 1). Both pH24R and pH24L were as active as ΦX174 RFI DNA in supporting ss DNA synthesis and in packaging newly synthesized ss DNA into proheads. The particles synthesized in these reactions were infectious to ΦX174-sensitive *E. coli* as judged by the transduction of tetracycline resistance (Tc^r) carried by pH24R and pH24L. The synthesis of ss DNA and infectious particles required the same components as for the ΦX174 RFI DNA-directed phage synthesis, that is, template DNA, purified ΦX174 gene A protein, gene C protein, gene J protein, prohead and uninfected host cell extract. Neither pH13 DNA nor pACYC184 DNA could support synthesis or packaging of ss DNA, indicating that the cloned ΦX174 fragment containing the A-site sequence carried necessary and sufficient information for ss DNA synthesis and packaging of the ss DNA into particles.

0028-0836/82/250704—03$01.00

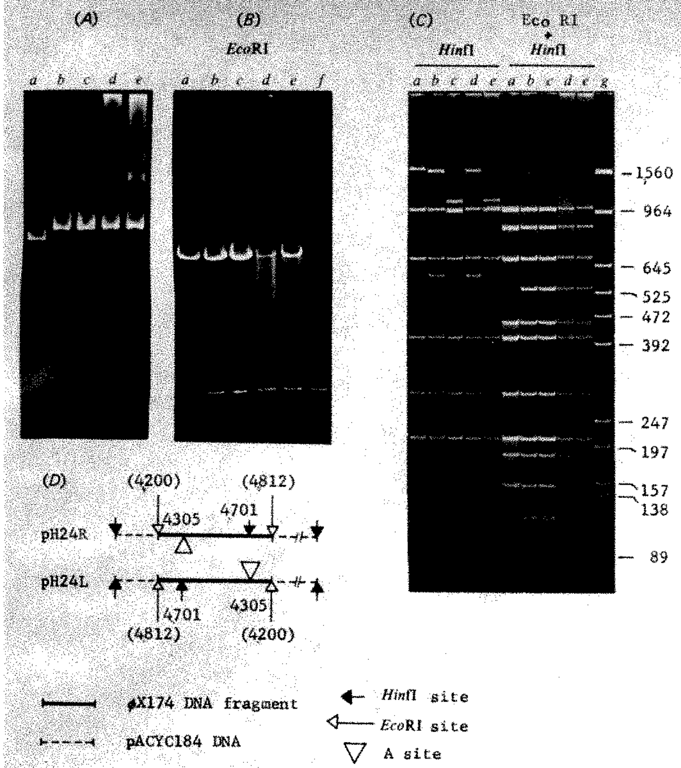


Fig. 1 Analyses of transduced DNA by gel electrophoresis. The *in vitro* reaction and the infection of *E. coli* HF4704 with the product were performed as described in Table 1 using either pH24R or pH24L DNA as template. DNA was isolated from Tc-resistant cells by the method of Birnboim and Doly[9]. DNA samples examined are: *a*, pACYC184; *b*, pH24R; *c*, pH24L; *d*, DNA extracted from Tc-resistant cells infected with *in vitro* particles synthesized in the pH24R DNA-directed reaction; *e*, DNA extracted from Tc-resistant cells infected with *in vitro* particles synthesized in the pH24L DNA-directed reaction. *A*, approximately 0.1 μg of each DNA was electrophoresed in a 1% agarose slab gel in Tris-borate buffer[12] (10.8g trisma base, 0.93g EDTA and 5.5g boric acid per litre) at 25 mA for 2 h. *B*, approximately 0.2 μg of each DNA was digested with endonuclease *Eco*RI and analysed as described in *A*. Fragment III of *Hin*cII digest of ΦX174 RF DNA which contains the A-site was electrophoresed as a marker (panel *f*). *C*, approximately 1.5 μg of each DNA was digested with endonuclease *Hin*fI or 4 μg of each DNA with endonucleases *Hin*fI and *Eco*RI. Gel electrophoresis was performed in a 7% polyacrylamide (acrylamide:methylene bisacrylamide = 29.2:0.8) slab gel containing 5% (v/v) glycerol in Tris-borate buffer at 10 mA for 16.5 h. ΦX174 RF DNA digested with endonuclease *Rsa*I was electrophoresed as markers for molecular weight (panel g) and the number of base-pair of each fragment is shown. *D*, Schematic representation of the structure of pH24R and pH24L DNA. Each number shown is the position of Sanger *et al.*'s DNA sequence of ΦX174 genome. *Hin*cII sites of ΦX174 RF DNA (4,200 and 4,812 of the Sanger *et al.*'s map) were attached to *Eco*RI linkers and is represented as *Eco*RI sites. Gels were photographed under UV light after ethidium bromide staining.

DNAs in the particles synthesized *in vitro* directed by pH24R and pH24L (pH24R* and pH24L*) were complementary as judged by the hybridization experiments (Table 2). Thus the orientation of the A-site sequence in the plasmid determines which strand of the plasmid DNA is synthesized and packaged into particles. Figure 1*A* shows that plasmid DNA isolated from Tc^r cells after infection with the *in vitro* particles (transduced DNA) co-migrated with the template DNA upon agarose gel electrophoresis. Fragments produced by endonuclease *Eco*RI (Fig. 1*B*) and *Hin*fI (Fig. 1*C*) digestion of transduced DNA also comigrated with their respective template DNAs. The *Hin*fI digest also shows that the template DNAs (pH24R and pH24L) have the A-site fragment inserted in opposite orientation (Fig. 1*D*). The pattern produced by digestion with *Hin*fI and *Eco*RI shows that the A-site sequence is inserted at the *Eco*RI site of pACYC184 (Fig. 1*C*, *D*). The recovery of the transduced DNAs as ds plasmid DNA indicates that the ss

Table 1 Template specificity and requirement for *in vitro* phage synthesizing system

Template DNA	Omission	ss DNA synthesis	Infectivity
ΦX174 RFI	None	148	7×10^{10}
pH24R	None	98	5×10^{10}
	A protein	5	6×10^{8}
	C protein	2	1×10^{8}
	J protein	7	$<10^{3}$
	Prohead	<1	$<10^{3}$
	Host protein fraction	<1	7×10^{4}
pH24L	None	90	8×10^{9}
	A protein	6	3×10^{8}
	C protein	4	4×10^{7}
	J protein	18	$<10^{3}$
	Prohead	<1	$<10^{3}$
	Host protein fraction	<1	3×10^{4}
pH13	None	<1	$<10^{3}$
pACYC184	None	<1	$<10^{3}$
None	None	<1	$<10^{3}$

Purified ΦX174 gene A protein, ΦX174 gene C protein, ΦX174 gene J protein, ΦX174 prohead, ΦX174 RFI DNA and *E. coli* protein fraction were prepared as previously described[1]. Plasmid DNAs used as templates were isolated by the method of Burnboim and Doly[9] and further purified by methylated albumin column chromatography[10]. *In vitro* reactions using 1 pmole of ΦX174 RFI DNA (molecular weight 3.4×10^6) or 1 pmole of plasmid pH24R (3.04×10^6 daltons), pH24L (MW3.04×10^6), pH13 (MW3.20×10^6), constructed *in vitro* as described in the text, or pACYC184 (MW2.65×10^6) were performed in the phage-synthesizing reaction mixture (25 μl) as described at 30° C for 30 min. The ss DNA synthesis is represented as pmole of deoxyribonucleotides incorporated into acid-insoluble, DNase-resistant materials in a 25-μl reaction mixture. DNase-resistant DNA is the DNA that has been packaged into proheads[1]. The infectivity of the product directed by ΦX174 RFI DNA as template was determined by plaque assay using *E. coli* HF4704 (sup°) as indicator bacteria and is represented as plaque-forming units (PFU) per ml of reaction mixture. The infectivity of the products directed by plasmid DNA was determined by infecting HF4704 with dilutions of the reaction mixtures and spreading on LB-agar plates (10g tryptone, 5g yeast extract, 10g NaCl and 15g agar per l) containing 20 μg ml^{-1} tetracycline. The infectivity is represented as the number of Tc-resistant cells produced per ml of reaction mixture.

Table 2 Hybridization experiment of *in vitro* packaged DNA

DNA source	c.p.m. hybridized
pH24R* + pH24R*	185
pH24R* + pH24L*	2,322
pH24L* + pH24L*	180

In vitro phage synthesizing reactions directed by either pH24R or pH24L DNA were performed as described in the legend to Table 1 except that the volume of the reaction mixture was 250 μl. After 30 min of reaction, DNase I was added to a final concentration of 0.1 mg ml^{-1} and the mixture was incubated for another 30 min. Pronase and sodium dodecylsulphate were then added to final concentrations of 0.5 mg ml^{-1} and 0.1% respectively and the mixture was incubated at 37 °C for 3 h. DNA was extracted by the hot-phenol procedure[11]. Each DNA sample (3,250 c.p.m.) was mixed and the mixture was adjusted to 35% formamide and 0.2 M sodium acetate (final volume 40 μl). After incubating at 35 °C for 13 h, 160 μl of water and 20 μl of S_1 nuclease solution (40,000 units ml^{-1} single-stranded specific nuclease S_1, 30 mM sodium acetate, *p*H 4.6 and 5 mM $ZnCl_2$) were added and the mixture was incubated at 37 °C for 1 h. Ice-cold trichloroacetic acid (7%)-precipitable counts were taken as hybridized counts. pH24R*, DNA extracted from particles synthesized in pH24R DNA-directed reaction; pH24L*, DNA extracted from particles synthesized in pH24L DNA-directed reaction.

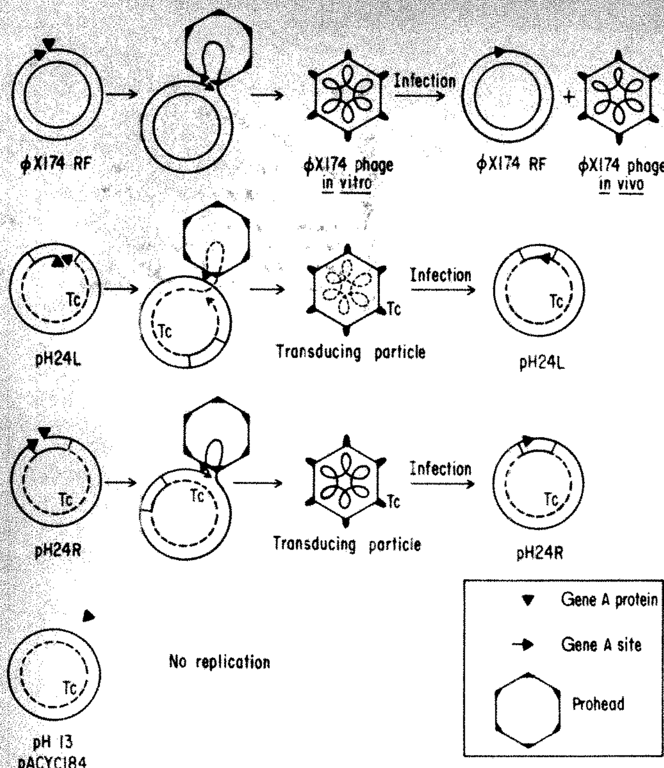


Fig. 2 Schematic representation of synthesis of infectious transducing phage *in vitro*. In the phage synthesizing system *in vitro*[1], ΦX174 gene A protein cleaves the viral strand of ΦX174 RFI DNA at its origin of replication (A-site) to form open circular RF (RFII) DNA whose 5'-terminus is covalently attached to gene A protein[4,5]. The RFII DNA-gene A protein complex then associates with prohead and ss DNA synthesis occurs by a looped rolling-circle mechanism[4]. Host proteins and gene C, gene J proteins of ΦX174 are required in these processes[1] (not shown in figures). The displaced viral strand is encapsidated into prohead while DNA synthesis is occurring. When one round of DNA synthesis is completed, gene A protein cleaves the viral strand and joins the two ends to form a phage particle containing a circular viral strand[6]. Phage particles synthesized *in vitro* are infectious to ΦX174-sensitive *E. coli* cells. After infection, the circular ss DNA is converted to ds RF DNA, which is multiplied to produce progeny RF molecules. In the late stage of infection, progeny RF molecules serve as template for synthesis and packaging of ss DNA to produce phage particles by a similar mechanism as *in vitro* reaction. In the pH24R or pH24L DNA-directed *in vitro* reaction, ΦX174 gene A protein cleaves either one of the strands of the plasmid DNA at the A-site and forms a plasmid DNA-gene A protein complex. After associating with prohead, asymmetric ss DNA synthesis occurs as in the case of the ΦX174 DNA-directed reaction. Depending on the orientation of the A-site fragment, one of the strands of plasmid DNA is specifically displaced and encapsidated into prohead. Thus pH24R or pH24L DNA-directed reaction produces infectious particles in which complementary circular plasmid ss DNAs are packaged. After infecting ΦX174-sensitive *E. coli.* cells, the plasmid ss DNA is converted to ds plasmid DNA, which is multiplied and maintained in the cell. As a result, tetracycline-resistancy carried by pH24R and pH24L is transformed to *E. coli.* cells. Neither pH13 DNA nor pACYC184 DNA support synthesis or packaging of DNA due to the lack of the A-site sequence in these plasmid DNAs.

DNA in the transducing particle is converted to and maintained as the ds form in the cell. The maintenance of the transduced plasmid probably occurs via the plasmid replication pathway because it depended on the presence of polA1 protein (data not shown).

The synthesis and packaging of ss DNA from pH24R and pH24L to produce transducing particles also occurred *in vivo* when cells harbouring these plasmids are infected with wild type ΦX174. The infecting phage apparently provides the necessary phage-originated proteins to produce the transducing particles by a mechanism similar to that occurring in the *in vitro* system described in this report. A schematic representation of the *in vitro* reactions is shown in Fig. 2.

This system will be useful in elucidating the biochemical functions of gene C and gene J proteins. By varying the size of the plasmid carrying the A-site sequence, this system also provides an excellent way to determine the minimal and maximal lengths of ss DNA able to be packaged into the prohead structure—an important parameter for the analysis of phage morphogenesis. The system described in this report also provides some applications in the molecular cloning of DNA. The efficiency of infection of particles containing plasmid DNA (as measured by the transduction of the drug marker) is as high as the efficiency of plating of phage, thereby providing several magnitudes higher transfer efficiency of plasmid DNA into cells over the conventional transformation method. The strand separation of plasmid DNA resulting from its packaging may also be useful in DNA sequencing.

We thank R. K. Hamatake for critical reading of this manuscript. This work was supported by NIH grant GM12934 and NSF grant PCM8011741 to M.H.

Received 21 December 1981; accepted 5 April 1982.

1. Aoyama, A., Hamatake, R. K. & Hayashi, M. *Proc. natn. Acad. Sci. U.S.A.* **78,** 7285–7289 (1981).
2. Dressler, D., Hourcade, D., Koths, K. & Sims, J. in *The Single-Stranded DNA Phages* (eds Denhardt, D. T., Dressler, D. & Ray, D. S.) 187–214 (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1978).
3. Hayashi, M. in *The Single-Stranded DNA Phages* (eds Denhardt, D. T., Dressler, D. & Ray, D. S.) 531–547 (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1978).
4. Eisenberg, S., Griffith, J. & Kornberg, A. *Proc. natn. Acad. Sci. U.S.A.* **74,** 3198–3202 (1977).
5. Langeveld, S. A., van Mansfield, A. D. M., Baas, P. D., Jansz, H. S., van Arkel, G. A. & Weisbeek, P. J. *Nature* **271,** 417–420 (1978).
6. Fujisawa, H. & Hayashi, M. *J. Virol.* **19,** 417–424 (1976).
7. Chang, A. C. Y. & Cohen, S. N. *J. Bact.* **134,** 1141–1156 (1978).
8. Sanger, F. *et al. Nature* **265,** 687–695 (1977).
9. Birnboim, H. C. & Doly, J. *Nucleic Acids Research* **7,** 1513–1523 (1979).
10. Mandell, J. D. & Hershey, A. D. *Analyt. Biochem.* **1,** 66–77 (1960).
11. Guthrie, G. D. & Sinsheimer, R. L. *Biochim. biophys. Acta* **72,** 290–297 (1963).
12. Davis, R. W., Botstein, D. & Roth, J. R. in *Advanced Bacterial Genetics,* 148–152 (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1980).

Errata

In the letter 'A 2-D model calculation of atmospheric lifetimes for N_2O, CFC-11 and CFC-12' by M. K. W. Ko and N. D. Sze, *Nature* **297**, 317–319 (1982), figures 2 and 3 have been transposed.

In the letter 'Helium isotopic systematics of oceanic islands and mantle heterogeneity' by M. D. Kurz *et al. Nature* **279,** 43–47 (1982), on page 45 the first two lines (with age ... degassed of ^{3}He) should be transposed to the bottom of the column.

Corrigenda

In the letter 'Structural alterations in J regions of mouse immunoglobulin λ genes are associated with differential gene expression' by J. Miller, E. Selsing and U. Storb, *Nature* **295,** 428–430 (1982), the sequence of the dJλ2 recognition nanomer is incorrect as shown in Fig. 2. The sequence should read GGATC<u>TT</u>GC. This change (underlined) only results in a closer resemblance of the dJλ2 recognition sequence to the consensus, making dJλ2 more like ψJλ3 and, therefore, more likely to interfere with functional Vλ2–Cλ2 joining. Thus, the error does not affect the conclusions presented.

In the letter 'Electrophysiology of mammalian thalamic neurones *in vitro*' by R. Llinas and H. Jahnsen, *Nature* **297,** 406–408 (1982), in line 4 in the right-hand column on page 407 the term 'early Na conductance' should read 'early K conductance'.

BOOK REVIEWS

Porcupine biology

A. J. Cain

THIS solid book (almost 900 pages in length) is essential reading for everyone at all interested in evolution, in biology or its history, or in science in general. Ernst Mayr is of course a contributor to the modern theory of evolution of top international standing, which instantly gives him a status far above most of those who have written on the history of biology; but also, he is a member by birth of one great European culture and a participant by adoption in the vast North American ethos. With such insights as are given by multilingual ability, breadth of culture, and a great practical experience and successful achievement in evolutionary research, one might expect him to produce a far more balanced, penetrating and illuminating book than could most others. Taking the book as a whole, it comes well up to expectation.

Mayr is as concerned with the conceptual structure of biology and its place among the sciences as with the development of particular themes within it. The introduction on how to write the history of biology, and the following chapter, "The Place of Biology in the Sciences and Its Conceptual Structure", develop his general methodology and viewpoint on the nature of biology; and the epilogue, "Towards a Science of Science", sums up and justifies much of his approach in the light of the findings in the main part. This latter falls into three parts: "Diversity of Life", "Evolution" and "Variation and Its Inheritance". Each of these is a fascinating history, from earliest times virtually up to the present, of taxonomy and systematics, of the idea of evolution, and of the various theories of inheritance up to the modern discovery of the diversity of DNA. The prose is usually excellent, though in a few respects American misusages have been appropriated by Mayr rather too thoroughly. Not surprisingly, American has some of the characteristics of a pidgin English, including the inability to distinguish between words of similar sound, however important and useful their distinctions of meaning; and if the proofs I have worked from are any guide, there is a near-total confusion between alternate and alternative, masterful and masterly, etc. Prepositions also suffer somewhat ("grist *on* the mill"). Equally American is his sur-

The Growth of Biological Thought: Diversity, Evolution, and Inheritance. By Ernst Mayr. Pp.896. ISBN 0-674-36445-7. (Harvard University Press: 1982.) $30, £21.

prise over Darwin's procrastination — "one would think that he would rush this, the most important theory in biology, to the printer as quickly as possible", (p.420). A peculiarly Mayrian trick is constantly to put words in inverted commas (sometimes totally unnecessarily) to warn the reader that he does not mean quite that. But what

exactly then does he mean? There is no place for inverted commas in scientific prose except to mark quotations.

Nevertheless, this is criticism at a high level. There is no dullness in the book and very little obscurity, but for whom was it written? Mayr indicates (p.19) that

> it is not technical to the extent that a layperson would have difficulty with the exposition. It is a major advantage of the history of ideas in biology that one can study it without a background knowledge of the name of a single species of animal or plant or of the major taxonomic groups and their classification. However, a student of the history of ideas must acquire some knowledge of the dominant concepts in biology . . . .

One can also read *Hamlet* omitting the Prince (one of Darwin's daughters wanted to omit Falstaff from *Henry IV*). The history of biology without a knowledge of the material generating and checking the concepts is precisely that useless and emasculated history of to-ing and fro-ing of concepts, so tiresome and uninformative, which is produced when historians and philosophers write histories of biology. At that rate science becomes as bad as philosophy. In science one must not only generate new concepts; they must be shown to be applicable to the real world.

Fortunately Mayr does not act on this ruinous precept. One of the virtues of the work is that it does cope, though not always sufficiently, with what controls that part of biology (or any other science) which really is science, namely the actual materials and phenomena investigated. A glossary of four whole pages is inserted to help the "layperson", but when one finds that mastodon is in it, but not stochastic, *Fragestellung*, deme, ichthyosaur, phenetic, euphorbs, chorda, *Bauplan* or *Urpflanze*, and that quinarian appears on p.102 but is not explained until p.202, Down's syndrome on p.581 with explanation only on p.759, neither being in the glossary, it becomes evident that Mayr is really writing for Simpson, Dobzhansky, Huxley and other professional students. This is not surprising. It is extremely difficult to abandon one's habitual universe of discourse.

Moreover, it is a very good thing that Mayr has not done so. He remarks himself that historians are not always aware of how complex biological concepts are. It is one of the great strengths of his book that he frequently gives tabulated or listed components of particular theories or concepts, in order to disentangle the numerous themes involved in them. I for one would be delighted if this book serves to put off a whole lot of intellectuals intending to burst into print on evolution and related topics.

In his exposition of his approach, Mayr includes an amusing personal apologia (pp.9–10):

> Whenever possible, I have attempted a synthesis of opposing viewpoints (unless one of them is clearly in error). Where the situation is quite unresolved, I have described the opposing viewpoints in categorical, sometimes almost one-

sided, terms in order to provoke a rejoinder, if such is justified. Because I hate beating about the bush, I have sometimes been called dogmatic . . . [but] This has never been my attitude, and, indeed, I pride myself on having changed my mind on frequent occasions.

When a man gets round to saying such things of himself, his colleagues have usually been saying them for the last 30 years, as I can testify personally; I can also testify personally to Mayr's changes of mind (and his candour in notifying them in print). Those Americans who in spite of their rugged frontiersman images have not had the guts to stand up to him have only themselves to blame. Nevertheless, even Mayr admits that this strategy is debatable. It all depends on whether it generates more heat than light, and a good deal of avoidable heat is going to be produced by critics of the book. Still, it must be admitted that a clear formulation of wrong ideas can be a real help in the advancement of science, as with Lyell (p.495). As Francis Bacon said, we can learn from error, not from confusion. Wynne Edwards has done similar service in his discussion of group selection, as has Popper in the methodology of science. (Mayr does not notice their subsequent retractions.)

Mayr is fully aware — how could he not be? — of his own magistral position in the modern synthesis, and some references to himself, e.g. p.449, are not quite as modest as are his remarks about his own limitations (e.g. pp.7, 400). The limitations a man is aware of, he can do something about; it is those he is unaware of that are serious. Three seem to me particularly notable. One is his lack of experience in handling plants. The book is a zoologist's book, even where plants are mentioned. Some of his harsher remarks about de Vries could not have been written by a botanist, a horticulturist or even a gardener. The varieties one actually finds in plants *are* either fluctuations or modifications, until a good deal of work has been done on them. "So-called 'difficult' genera like *Rubus* or *Crataegus*" (p.263), *are* difficult, and the difficulty cannot be surmounted by setting them aside. Similarly Mayr fails to appreciate the strength of botanists' objections to Weismann (which affected some zoologists as well), and his references to polyploidy and allo-tetraploidy in plants (pp.361, 645) are as perfunctory as in his contributions to *The Evolutionary Synthesis* (Harvard University Press, 1980). To call polyploidy a special case in speciation is to provoke botanists in quite the wrong way. They might well ask whether birds and insects are not special since they manifest so little of it. Equally, single-gene bases for sexual isolation have been known for a long time in plants; that only a small part of the genotype is concerned is not such a discovery as Mayr seems to think (p.605).

Secondly, he is not particularly at home among the invertebrates. This matters much less than his neglect of the botanical side in this book, but occasionally shows as when he speaks of Lamarck (pp.346, 367) as establishing on his fossil shells "virtually unbroken phyletic series". To anyone who has worked through Lamarck's species in the *Histoire des Animaux sans Vertèbres* this is a misleading exaggeration.

Thirdly, and perhaps most important, he has never worked on natural selection, only on the consequences of it. He misses throughout the book one of the most extraordinary facts about work on evolution, namely that hardly anyone has done anything on natural selection in the wild. Almost every worker has *talked* about it often at a great length and with subtle mathematics, but very few have had any experience of it actually in action and much of the recent talking has been little better than politico-religious propaganda. If indeed one wishes to confine oneself to concepts, one can go on treating the questions raised virtually as ideological, which also can generate heat, but hardly light. The treatment of the actual role of natural selection, as far as we can understand it up to the present, is one of the weaker features of the book.

But having said this, again at a high level, one must also say that there is no other book like it for range, depth and insight into the nature and development of these three major biological themes, and the methodology and position of biology among the sciences. Particularly good are Mayr's analyses of essentialism and typological thinking as against population thinking, of the differences of interest and approach between those who are fascinated by diversity and by adaptation, of the necessity for both historical types of explanation and immediate causal ones in biology, and of the incompatibility of the approach of experimentalists and naturalists. (One should now add a third category, the computer-user, with knowledge of organisms neither in the wild nor in the laboratory.) By the side of this book, Singer is trivial, Radl a hopeless *a priorist*, Nordenskiold a religious propagandist.

One could of course list a considerable number of minor, or not-so-minor errors; some are almost inevitable in a book of this scope, but I will confine myself to four, answering Mayr in his own style. (1) p.25: "The Greeks always looked for rational explanations in the world of phenomena". This is pure typological thinking; most did not, and of the few that did, some had a pretty bad time because of it. Read E.R. Dodds's superb *The Greeks and the Irrational* (University of California Press, 1951). (2) pp.361–362: "The Enlightenment ended, so to speak, with the French Revolution . . .". It did not; look at the followers of Paine and Bentham, and Robert Chambers. (And what does "so to speak" mean?) (3) p.510: Huxley has a perfectly explicit reference to natural selection and several indirect ones in his

essay on the reception of the *Origin of Species*; to say that he has none is wrong, and to label him as an essentialist is again typological thinking. And lastly, a more general point (4) pp.375, 504 etc.: A better appreciation of natural theologians and scientists in England is achieved if one considers the reliability of *testimony*; in science, there is one's own experience, the personal witness of acquaintances, and the testimony of others in papers and books. One is unique in this respect in being the testimony of the Author of Creation, and therefore particularly valuable.

Certain classes of mind, Marxist in particular, will detest this book; most others will find it exceedingly stimulating, not to say infuriating. It has as many points as a porcupine, for intellectual aggression, not defence. There will be a lot of love–hate relationships engendered. I love it, spikes and all. □

A.J. Cain is Derby Professor of Zoology at the University of Liverpool.

High time for unification in catalysis

R.O.C. Norman

Metal-catalyzed Oxidations of Organic Compounds. By Roger A. Sheldon and Jay K. Kochi. Pp.424. ISBN 0-12-639380-X. (Academic: 1981.) $56, £37.

METAL-catalysed oxidations of organic compounds have been attracting increasing attention over the past decade or so in three conventionally distinct fields of chemistry: liquid-phase oxidations, traditionally an area of study by the organic chemist; heterogeneous gas-phase reactions, the province of the physical chemist; and enzymatic reactions, that of the biochemist. Many of us who are involved in one of these branches of the subject have sometimes been slow to appreciate the value to our own field of an advance in either of the other two, and it was high time someone attempted to unify the subject. Moreover, for obvious reasons, metal-catalysed oxidations have strong attractions to industrial chemists, who can almost certainly be helped in their task by an understanding of how nature works; for example, it is probably not an exaggeration to suggest that a full understanding of how cytochrome P450 acts in the living cell could help an industrial project aimed at catalysing the aerial oxidation of benzene to phenol.

In this book the task of unification has been undertaken by two men who are well suited to it by virtue of both their complementary fields of experience — industry and academia — and their distinguished activities as researchers in the field. They have succeeded admirably.

The structure they have adopted is logical: to consider first the mechanistic principles of the oxidations, grouped by reaction type — oxidations by molecular oxygen, by peroxides, by oxometal reagents and so on; and then to describe synthetic methodology, grouped by compound type — olefins, aromatic hydrocarbons, alkanes, oxygen-containing compounds, and nitrogen, sulphur and phosphorus compounds. It is true that this results in a certain amount of repetition; for example, the mechanism of oxidation of arenes by cobalt(III) on p.122 is repeated on p.319. This may irritate some readers but, by and large, it serves to reinforce one's knowledge and understanding in proceeding through a complex subject.

The coverage is enormously comprehensive, including details of reactions catalysed by (salts of) 33 metals, with well over a thousand references, including some from 1981. To some extent, then, the account reads more like a review than a textbook. The disadvantage of this is that little, almost trivial, points are occasionally included for completeness and slightly upset the flow and development of the major themes; but this is certainly outweighed by the advantage of having such a valuable collection of well-referenced information so readily to hand.

R.O.C. Norman is Professor of Chemistry at the University of York.

Moving pictures

P.W. Hawkes

Image Sequence Analysis. Edited by T.S. Huang. Pp.437. ISBN 3-540-10919-6; 0-387-10919-6. (Springer-Verlag: 1981.) DM85, $39.60.

PROCESSING a single black-and-white image in a computer requires a substantial amount of memory and, all too often, of time. If we add colour, the difficulties become appreciably but not insuperably worse. As soon as we are driven to try and extract information from large numbers of images, though, the complexity and immensity of the problem become such that all but the very best-equipped must surely blench. Reading between the lines of this interesting and detailed discussion of the problems of extracting information about moving objects from image sequences, and related topics, I cannot help feeling that Professor Huang and his colleagues must occasionally have quailed; nonetheless the overall impression is one of optimism.

The book is divided into three parts. The first consists of an introduction, largely devoted to the estimation of motion parameters, and a review by H.H. Nagel of the various applications (with an extensive bibliography and an author index). The second, on image sequence coding, enhancement and segmentation, contains four chapters: "Image Sequence Coding" by E. Dubois *et al.*; "Image Sequence Enhancement" (the editor and Y.P. Hsu); "Image Region Extraction of Moving Objects" (B.M. Radig); and "Analysing Dynamic Scenes Containing Multiple Moving Objects" (J.K. Aggarwal and W.N. Martin). The final section, on medical applications, consists of a single chapter on processing medical image sequences by W. Spiesberger and M. Tasto.

The possible applications of this work are, in many cases, of considerable industrial importance, with the result that although much abstract and abstruse material is presented, we are repeatedly brought back down to earth by the specific illustrations. In the introduction, Huang and Tsai offer the following by no means exhaustive list of potential applications: road traffic monitoring; cloud tracking; microcinematography and X-ray sequences of moving parts of the body, the heart in particular; bandwidth compression of picture-phone and TV conferencing signals; robot vision and dynamic monitoring in industry; and, inevitably, target tracking for military purposes. Readers from other fields will have no difficulty in extending this list: *in situ* studies in electron microscopy, the behaviour of nerves and muscles — there can be few fields where studies of rates and types of change could not benefit from some automation.

Most of the book is concerned with image coding and with the study of rates of change of position, ranging from the position of cars or missiles to the position of a contrast medium in a diseased and a normal kidney. It is written in such a way that the newcomer familiar with "static" image processing can understand in detail the methods used to analyse the "dynamic" situation and, at the same time, the absolute beginner — industrialist, tycoon, general, doctor — can comprehend at least the kinds of problems for which the methods have been devised and judge whether they will be of any help to him. In short, this is a meaty, fairly readable and eminently useful addition to the image processing literature. □

P.W. Hawkes is Maître de Recherches at the Laboratoire d'Optique Electronique du CNRS, Toulouse.

Mineralogy for everyman?

B.E. Leake

The Encyclopedia of Mineralogy. Encyclopedia of Earth Sciences, Vol. IVB. Edited by Keith Frye. Pp.794. ISBN 0-87933-184-4. (Hutchinson Ross: 1982.) $95, £62.80.

WHAT does the non-mineralogist want to know about mineralogy? This encyclopaedia sets out to provide the answers but of course those answers are determined by what the editor conceives the reader will want to know, and that is difficult to predict.

According to the preface, the book is mainly intended for the non-mineralogist and the advanced professional is expected to "have little use for our volume". The scope is therefore wide, and the references to mineralogical journals and museums are extensive so that inquirers can follow topics to a greater depth than a general survey can provide. Just over 100 writers — most of them in the United States — contribute 140 articles, ranging alphabetically from abrasive materials to zeolites. These are followed by a mineral glossary containing about 3,000 entries of mineral groups, species and varieties, including mineraloids, in which brief descriptions of the chemical formula, physical properties, occurrence and use of each mineral are terminated by references to relevant literature.

In any encyclopaedia the key element is the index. However well organized, erudite or comprehensive the coverage, the information extractable from such a book is controlled by the quality of the index. For mineral names the index is good, but it is inadequate for many other topics on which would-be users could well consult the book. The mineralogical composition of common building bricks or porcelains? Not even the word ceramic appears in the index. Which crystals are used in X-ray fluorescence analysis? What is diffraction? None of these questions can be answered by consulting the index; some of them no doubt lie in other volumes of this mammoth series of earth science encyclopaedias, of which this is only Vol. IVB. Moreover any reference book must have omissions vulnerable to criticism. The mineralogies of cements and glasses are described, however, so the coverage is not restricted to natural occurrences.

In my view the book is too technical for the non-mineralogist. It is three-quarters designed for the professional mineralogist — who will normally prefer more extensive texts or *Mineralogical Abstracts* — although the mineral glossary, which occupies a quarter of the volume, will be useful as a quick guide into the literature. The information provided about most minerals is that which a *mineralogist* would normally be concerned with, such as optical and crystallographic properties. Items such as chemical properties (as distinct from chemical composition) are rarely considered so that the solubility of minerals in water, the effects of heating minerals together etc. are scarcely dealt with, although eight pages are accorded to the blowpipe analysis of minerals and five to staining techniques, both of which are clearly described. No doubt a major problem for the editor was the lack of available information of an unconventional type, since most data on minerals are provided by mineralogists and the articles were thus largely written by them.

Nonetheless if you want to know about piezoelectricity, phantom crystals, geological barometry and thermometry, pleochroic haloes and thermoluminescence, it's there in a potted form. Also Epsom salt and human and vertebrate minerals, but unless you are a crystallographer the article on point groups and the enumeration of the twin laws of the feldspars will surely leave you baffled — the list of the Council members of the International Mineralogical Association from 1958 to 1978 is more readily understood.

The book, then, will be a useful addition to general libraries. But it is too complex for the non-mineralogist and largely repeats information already available in numerous, recently published mineralogical textbooks. □

B.E. Leake is the Professor of Geology at the University of Glasgow, and Honorary Keeper of the Geological Collections in the University's Hunterian Museum.

The second volume of the multi-volume treatise *Social Insects*, edited by Henry R. Hermann, has recently been published by Academic Press. Price is $55, £36.40. (For a review of Vol.I see *Nature* **282**, 884; 1979.)

Sex, parthenogenesis and the tangled bank

Mark Ridley

The Masterpiece of Nature. By Graham Bell. Pp.635. UK ISBN 0-85664-753-5; US ISBN 0-520-04583-1. (Croom-Helm/University of California Press: 1982.) £25, $45.

"SEXUAL generation", wrote Erasmus Darwin, "seems the chef d'oeuvre, the master-piece of nature". Sex is also one of biology's master-problems, and has been honoured by three of the master-works on evolution which have appeared in the last decade: M.T. Ghiselin's *The Economy of Nature and the Evolution of Sex* (California University Press, 1974), G.C. Williams's *Sex and Evolution* (Princeton University Press, 1975) and J. Maynard Smith's *The Evolution of Sex* (Cambridge University Press, 1978). Graham Bell has combined the theorizing and algebra of a Williams and a Maynard Smith, with the comparative biology of a Ghiselin, to produce a book which is, indeed, more than the length of the other three put together. *The Masterpiece of Nature* is a pleasantly written and important work; it can stand with the other three on the shelf, even if it does not edge them over into the waste-basket as its conclusions sometimes imply.

First, some praise for the 170-page review of sexuality which comprises Chapter 3. It is an immense achievement, which systematically reviews the literature on sex and parthenogenesis for all the taxa of multicellular animals. Despite a few minor omissions, it will be of great value as the only available zoological encyclopaedia of sex.

Sex is a puzzle because it has an apparent two-fold disadvantage compared to asexual reproduction. A mutation which caused a female bearer to reproduce asexually would go into all its offspring rather than just half of them. And the offspring would all be females, which lay eggs, whereas half of the sexual female's offspring are males whose energetic contribution to the next generation is usually nothing. Recent work has mainly attempted to discover some unnoticed advantage to sex sufficiently great to outweigh its huge, two-fold cost.

Bell takes a different line. He is less concerned with thinking up some new theoretical advantage to sex than with finding out which of the existing theories best explains the incidence of sex and parthenogenesis in nature. The facts meet the theories in Chapter 4, and Bell's verdict is that "the problem originally set has been solved" (pp.391–392) in favour of a theory which he calls "the tangled bank". The tangled bank is, in its emphasis, a new theory. It proposes that there is spatial variation of the environment; each species exploits several niches in space. An asexual clone could take over one niche, but then sex may become advantageous. The asexual progeny must compete among themselves within their one, already saturated, niche. The sexual progeny are more diverse, and so are adapted to a greater diversity of niches. Parthenogenesis is, in fact, most common in empty patches and other circumstances of reduced competition.

The tangled bank is closely related to models of what Maynard Smith has called "sib competition". So I was surprised when, in his chapter on the theories, Bell lumps the previous sib competition models

along with some others as "best man" theories, and discusses them separately from the tangled bank. The distortion at this stage is clear, but it does not become damaging until the two resurface as competitors in Chapter 4. To make the sib competition (best man) model into an opponent, he strips it of spatial variation, of temporal variation in all but physical factors, and demands that if sex is an adaptation to changing environments then the organism ought to reproduce sexually *in* the changed environment. The sickly resultant theory he then bullies with "as rigorous a comparative scrutiny as we can devise" (p.359). Its tongue thus cut out he then scolds it for silence: he "receives no explanation" for one fact after another, and the rhetoric of refutation resounds through the chapter. At the extreme he finds "incompatible with the lottery version of the best man" (p.368) a study which seems to me to fit it rather well: "Hebert has concluded . . . that most populations are founded by only one or a very few ephippia [of *Daphnia*]", which looks very like the sib competition model with a single survivor in each patch. Yet Bell concludes that "this offers little scope for the lottery principle".

Bell has deceived himself by his dichotomy between tangled bank and sib competition (best man) theories. But let him not deceive us as well. We can sap him by colouring back in the caricatured "best man". Then we won't be fooled by all the remorseless "refutation" and "rejection". We can watch Bell's fascinating advocacy of his tangled bank, all the time seeing him make the strongest case ever that some kind of sib competition is the best explanation of the patterns of sex in nature. □

Mark Ridley is Junior Research Fellow of Oriel College, and a member of the Animal Behaviour Research Group, Oxford University.

By any other name

Keith Bell

The Cambridge Encyclopedia of Earth Sciences. Edited by David G. Smith. Pp.496. US ISBN 0-517-54370-2; UK ISBN 0-521-23900-1. (Crown/Cambridge University Press: 1982.) $34.95, £19.95.

THE word "encyclopaedia" is one of the most awesome in our language, conjuring up visions of dull, pedantic prose, intimidating collections of facts and a morass of unconnected topics. To me, such volumes were always best left alone. Therefore, it was with some trepidation that I started thumbing through the contents of *The Cambridge Encyclopedia of Earth Sciences* expecting the worse. This was not to be.

Between the covers rests an extensive collection of topics, clearly written, beautifully illustrated and surprisingly up to date. Do not be fooled by the title — the only encyclopaedic feature of this book is its breadth and scope. An encyclopaedia it is not. The book is sub-divided into six fairly broad, distinct parts covering 27 topics that start with the history of the earth sciences and end with the geology of the Solar System. Included are accounts of such diverse fields as remote sensing, the development of life on land and seismic reflection profiling. There are no snappy definitions or thumb-nail sketches of your favourite topic here — this volume is definitely not of the "let's see what it says about such and such a subject" variety. Presented instead is a carefully planned, well-thought-out, overview of the earth sciences. Written primarily by members of the Open University, the book seems to be a much modified and markedly improved version of the popular Open University textbook *Understanding the Earth,* first published in 1971.

Its unfortunate billing as an encyclopaedia implies that information can be readily extracted from the book, but neither its format nor content allow this; it may take some time and a little effort to pull together information relevant to one particular topic. This is especially true of mineral deposits; information on the subject is scattered throughout many sections of the book.

As an introduction to the earth sciences this volume is one of the most exciting publications to emerge in the past few years, one worth having for the diagrams alone — many in colour. It is a text that will probably include among its audience scientists outside the field, interested laymen and introductory geology students. Although not aimed at the professional geologist, so much useful information abounds in this one volume that even those actively working in the earth sciences will find something of interest. I would relish recommending this book as a first-year textbook in the earth sciences, but I can already hear the groans as the title is written up on the blackboard.

Nevertheless, someone has had the verve and imagination to pull together a great deal of information about an area of science that is changing at a dazzling speed. Such a massive undertaking must have appeared formidable to any editor, but to plan and create a volume like this in the short span of three years is really quite remarkable. Its welcoming clarity and freshness should do much to intrigue a host of readers about the earth sciences, in addition to providing a quick refresher course to those geologists who, for one reason or another, have fallen behind. But, remember, it is a book that flies under a false flag. □

Keith Bell is a Professor of Geology at Carleton University, Ottawa.

Ideas of desert life

J.L. Cloudsley-Thompson

Biology of Desert Invertebrates. By Clifford S. Crawford. Pp.314. ISBN 0-387-10807-6. (Springer-Verlag: 1982.) DM89, $41.50.

DURING the present century, the study of desert biology has passed through at least three phases. An initial expansion of interest led to an enhanced understanding of the desert biome (exemplified by the late Professor P.A. Buxton's pioneering work, *Animal Life in Deserts*, first published as early as 1923). This was followed in the years after the Second World War by a considerable amount of research on the adaptations, physiology and phenology of desert animals. The third phase began with an awakening of interest in ecosystems, and the establishment of the International Biological Programme in 1964. Less than ten years later, interest in the environment had grown to a point at which another generation of biologists was viewing desert biology from the perspective of the community and the ecosystem. The impetus for this approach came partly from the intellectual and financial backing of the IBP, and partly from major advances taking place in ecological theory and physiological methodology.

Cliff Crawford's valuable statement not only brings together much of the work of past and present investigators, but reflects his own conceptual bias — probing into what desert invertebrates do and how they do it — in dealing with a topic not easily manipulated. His book is organized in five parts: Deserts and Desert Invertebrates; Adaptations to Xeric Environments; Life-History Patterns; Invertebrate Communities; Invertebrates in Desert Ecosystems: Summary Remarks. An idea of the detail included in this synthetic treatment is indicated by the fact that there are no less than 33 pages of references. In the last analysis, despite the recent surge of facts and interpretations, so evident in the book under review, we are plainly still far from a comprehensive understanding of the roles of desert invertebrates. Are not some of their patterns of adaptation, life history and community interactions more apparent than real? Continued study will doubtless reveal the truth in such matters.

Only to the uninitiated do arid regions appear dull and lifeless. Their inhabitants may, indeed, be secretive and few in number, but the biological problems posed by their very existence in such harsh and inhospitable surroundings makes the desert an exciting environment for research. Crawford's vivid text and striking photographs will undoubtedly stimulate the interest of the ininitiated, and serve as a reference and source of new ideas for research workers. □

J.L. Cloudsley-Thompson is Professor of Zoology at Birkbeck College, University of London.

Cancer as deviant differentiation

D.G. Harnden

Cancer Biology. By Raymond W. Ruddon. Pp.344. Hbk ISBN 0-19-502942-9; pbk ISBN 0-19-502943-7. (Oxford University Press: 1982.) Hbk $21.95; pbk $14.95, £10.

IT IS relatively easy to describe the characteristics of cancer cells but rather more difficult to say which ones are critical for conferring malignancy. Similarly, it is easy to conclude that there must be an alteration in the genome; but whether the change is structural, regulatory or a combination of the two has always seemed rather harder to define. It comes as something of a culture shock, therefore, to find a book which not only assumes that it is the regulatory changes that are fundamental but makes this thesis sound uncomfortably convincing for an advocate of somatic mutation. To quote Dr Ruddon (p.55):

> What we need to know to understand carcinogenesis and to develop ways of preventing or curing cancer, then, is contained in the mechanisms of normal cellular differentiation.

It has been clear for some time that an understanding of differentiation could prove to be very important in some cancers, but we now have to reckon with the possibility that, for the majority of cancers, the principal irregularity is a deviation from, or an arrest at some point in, the pathway of development of the stem cell into a mature differentiated cell. The association between mutagenesis and carcinogenesis, however, is too strong to be unimportant; but it may be that the link lies in the structure and function of regulatory sequences, whose potential for controlling the balance between preparation for cellular proliferation and the production of specific products of the differentiated cell we are only just beginning to understand.

Dr Ruddon does not go out of his way to advocate the importance of regulatory phenomena. Indeed in later chapters h deals quite fully and fairly with the role o mutational events. Some might call tl overall presentation unbalanced, but found it an interesting and stimulatin approach in which the author's person views and areas of interest feature qui prominently. This makes it much mor compelling to read than a multi-autho book and makes the breadth of coverag quite remarkable. Little prior knowledg of cancer nor indeed of complex biologica techniques is assumed, and this makes it very good book for senior students and those entering research projects in this field — as well as refreshing for those more fixed in their ways. □

D.G. Harnden is Professor of Cancer Studies a the University of Birmingham, England.

Ecological theory meets mycological data

C.T. Ingold

The Fungal Community: Its Organization and Role in the Ecosystem. Edited by Donald T. Wicklow and George C. Carroll. Pp.855. ISBN 0-8247-6956-2. (Dekker: 1981.) SwFr. 280, $98.50.

IN RECENT years mycologists have become increasingly interested in the life processes of fungi in their natural environments, and, at the same time, there has been a growing awareness amongst ecologists of the importance of fungi in the major ecosystems. The mycologist has tended to focus attention on specific niches in which fungi are the dominant organisms; in particular much attention has been given to fungi in the soil. However, the outstanding need is for valid quantitative estimates of the contributions of fungi to the dynamic equilibrium of such major plant communities as forest, prairie or heath. The ecologist is concerned with all of the components of these communities and how they interact. By their very nature, the fungi are perhaps the most difficult to handle.

The Fungal Community, as Swift says in his foreword, is an evangelistic work. It has the worthy aim of stimulating mycologists and ecologists to concerted and integrated endeavour. No less than 52 authors, including some general ecologists, have contributed 41 chapters which are divided into eight parts: fungal ecology and ecological theory; fungal populations; organization of fungal communities; interactions among fungal populations; fungal community development; fungal productivity in ecosystems; fungi in nutrient re-cycling; and finally the aquatic hyphomycete community as an example of how mycologists have approached a specific ecological problem.

The individual chapters are of divers nature and quality. Some are excellen such as Wicklow on the coprophilou community; Emerson and Natvig or adaptation of fungi to stagnant waters Frankland on fungal succession; Starme on yeasts associated with cacti and spread by species of *Drosophila*; Lindeberg on nitrogen cycling in coniferous forests; and Suberkropp and Klug on degregation of leaves by aquatic hyphomycetes. However, a number of chapters are mediocre. In all, one is left with the feeling that if, instead of 52 authors, a quarter of that number under closer editorial control had contributed, a more valuable and influential work might have resulted.

Three chapters deal with lichens, but the value of considering them in this book is doubtful. As Pike, in his chapter on lichen biomass, remarks "the thallus of lichens bears more similarity to the body of a green plant than it does to the mycelium of most fungi".

Nonetheless, *The Fungal Community* will be read with interest not only by mainstream ecologists concerned with the whole ecosystem, but also by the growing number of mycologists who are taking an ecological approach that is at the same time both experimental and quantitative. (Incidentally it is to be hoped that, in adopting and adapting ecological methods, these mycologists will be at pains to avoid indulging in too much ecological jargon.) The path ahead in the study of the ecology of fungi will undoubtedly be beset with difficulty, but the production of this book is one step forward. □

C.T. Ingold is an Emeritus Professor of the University of London, and a former President of the British Mycological Society.